KEER2026 ACADEMIC PAPERS

PRESENTED AUGUST 11-12, 2026
UNIVERSITY OF OREGON, EUGENE, OR, USA
DOI: 10.7264/XGCK5480

REVIEW & EDITORIAL BOARD

KIERSTEN MUENCHINGER, UNIVERSITY OF OREGON, USA

ANITAWATI MOHD LOKMAN, UNIVERSITI TEKNOLOGI MARA, MALAYSIA

TOSHIMASA YAMANAKA, UNIVERSITY OF TSUKUBA, JAPAN

SUOMIYA BAO, UNIVERSITY OF TSUKUBA, JAPAN

KUOHSIANG CHEN, NATIONAL CHENG KUNG UNIVERSITY, TAIWAN

PIERRE LEVY, CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS, FRANCE

SHINICHI KOYAMA, UNIVERSITY OF TSUKUBA, JAPAN

MIZUKI NAKAJIMA, TOKYO DENKI UNIVERSITY, JAPAN

KUNIKO OTOMO, UNIVERSITY OF TSUKUBA, JAPAN

FAUZIAH REDZUAN, UNIVERSITI TEKNOLOGI MARA, MALAYSIA

NAZLINAH SHAARI, UNIVERSITI PUTRA MALAYSIA, MALAYSIA

TZU-WEI TSAI, NATIONAL TAICHUNG UNIVERSITY OF SCIENCE AND TECHNOLOGY, TAIWAN

RACHAEL VOLKER, UNIVERSITY OF OREGON, USA

MARCO MENG-CONG ZHENG, NATIONAL TAIPEI UNIVERSITY OF TECHNOLOGY, TAIWAN

A FREE-ENERGY FORMULATION OF THE TRIADIC VALUE MODEL FOR INNOVATION OF MEANING

ANALYSIS OF THE EFFECT OF PRECISIONS ON VALUE-GENERATION PROCESSES BASED ON A GAUSSIAN MODEL

Kensaku TANIYAMA[1], Shotaro KUSHI[1,2], Hideyoshi YANAGISAWA[1*]

[1] *The University of Tokyo, Japan*

[2] *NEW STANDARD Inc., Japan*

[*] *Corresponding author: hide@mech.t.u-tokyo.ac.jp*

ABSTRACT

Triadic Value Model (TVM) describes value generation as a dynamic process among intrinsic value, instrumental value, and meaningful value, formulated as Bayesian inference. Meaning Generation Theory (MGT), on the other hand, provides a practical framework for generating new meanings through contextual shifts, but mainly corresponds to only one process within TVM. This study aims to bridge the research gap between TVM and MGT by formulating all value-generation processes in TVM using the free energy principle. By mapping the processes of meaning assignment, value updating, and value embodiment onto information-theoretic quantities—namely Kullback–Leibler divergence, Bayesian surprise, and expected free energy—we analytically derive design-relevant properties of each process. Using Gaussian approximations, we clarify how the variances of prior beliefs, likelihoods, and action-conditioned beliefs influence these quantities. The results provide desing-relevant guidelines for innovation of meaning, showing how exploration and exploitation can be optimally balanced through the manipulation of precision. This framework extends TVM toward a practical methodology and provides a unified theoretical foundation for innovation of meaning grounded in Bayesian cognition and active inference.

Keywords: *innovation of meaning, triadic value model, free energy principle, active inference*

1 INTRODUCTION

In recent years, innovation of meaning (Verganti, 2009) has become a central topic in design research and practice, based on the recognition that the value of products and services depends not only on their functional performance but also on how they are interpreted by people. While conceptual discussions on innovation of meaning have accumulated, methodological challenges faced by designers remain unresolved. In particular, it is difficult to describe value generation based on meaning in a way that is both theoretically coherent and practically operable. Specifically, it remains challenging to address, within a single continuous framework, how

meanings are reinterpreted and how goals and offerings are subsequently updated and embodied as a result of that reinterpretation (Norman & Verganti, 2014; Kushi & Yanagisawa, 2024).

To address this issue, the Triadic Value Model (TVM) conceptualizes value generation as the interrelationship among intrinsic value (goals), instrumental value (means), and meaningful value (value arising from interpreting the relationship between goals and means), and describes value generation as three dynamic processes (Yanagisawa, 2025). These processes are: (1) the generation of instrumental value through the embodiment of intrinsic value as means, (2) the generation of meaningful value through assigning relationships between instrumental value and intrinsic value (goals), and (3) the generation of intrinsic value through legitimizing meaningful value and updating the goals themselves (legitimization).

Meaning Generation Theory (MGT), on the other hand, provides a practical perspective based on context shifts, whereby the interpretation—and thus the meaning—of the same object changes as its context changes, leading to the generation of new meanings (Kushi, Nakaji & Yanagisawa, 2024). For example, a candle may be interpreted as a source of lighting in one context, while in a context oriented toward relaxation it may be interpreted as a means of self-care. In this way, MGT offers a useful framework for addressing innovation of meaning in design practice.

However, a significant gap remains between TVM and MGT. While TVM provides an integrated account of value generation as a dynamic triadic process, MGT primarily corresponds to only one of the three processes in TVM—namely, the generation of meaningful value through meaning assignment. As a result, (i) the process of generating instrumental value through the embodiment of intrinsic value and (ii) the process of updating intrinsic value (goals) through the legitimization of meaningful value lack sufficiently clear and practically applicable methodologies. Moreover, even the process of meaning assignment itself requires more concrete descriptions that allow it to be treated not merely as an ex-post explanation, but as a set of manipulable variables within design practice.

To bridge this gap, this study formulates the entire set of value-generation processes in TVM based on the Free Energy Principle (FEP) (Friston et al., 2006; Friston, 2010), a unifying principle in cognitive neuroscience. The FEP posits that perception, learning, and action can be described as the minimization of variational free energy, with Bayesian inference emerging as a consequence of this minimization dynamics. By mapping each process in TVM onto this minimization framework, this study aims to represent designers' value-generation operations in forms that are analytically tractable and directly connectable to design guidelines.

Specifically, meaning assignment is conceptualized in terms of Kullback–Leibler divergence (KLD) as a component of free energy reduction, value updating is conceptualized as Bayesian surprise (BS), and value embodiment (the generation of means) is conceptualized as expected free energy within active inference. Furthermore, key parameters such as the variance or precision of prior beliefs, likelihoods, and action-conditioned beliefs are treated as design variables. In a design context, these parameters correspond to the degree of commitment to

existing goals (prior beliefs), the reliability and clarity of perceived cues from artifacts (likelihoods), and the uncertainty associated with anticipated outcomes of design actions such as prototyping (action-conditioned beliefs). By analyzing how these parameters influence each information measure, this study derives design-relevant guidelines concerning the balance between exploration and exploitation in innovation of meaning.

In summary, the contributions of this study are twofold: (1) it reconstructs TVM as a process-complete methodology that addresses all three value-generation processes, and (2) it positions innovation of meaning within a unified framework consistent with free energy principle and active inference, thereby deriving design-relevant guidelines with implications for design practice.

2 METHOD

2.1 Triadic value model

TVM is a theoretical framework that models the dynamic processes of value generation (Yanagisawa, 2025). TVM describes value generation in terms of Bayesian inference. Bayesian inference is a method for probabilistically inferring causal states from observed events. According to the Bayesian brain hypothesis, the brain perceives the world through Bayesian inference (Knill & Pouget, 2004). Bayesian inference can be expressed as follows.

$$p(s|o) = \frac{p(s)p(o|s)}{p(o)} \tag{1}$$

Here, s denotes a state and o denotes an observation. This model infers the state of the environment, which is the cause, from observations. $p(s)$ represents the prior distribution, $p(o|s)$ the likelihood, and $p(s|o)$ the posterior distribution. These correspond to the predicted state, the observed state, and the perceived state, respectively. According to the Bayesian brain hypothesis, the brain does not directly perceive observations; rather, it perceives the product of prediction and observation.

In TVM, the prior distribution represents intrinsic value, the likelihood represents instrumental value, and the posterior distribution represents meaningful value. intrinsic value corresponds to the goal, instrumental value to the means for achieving the goal, and meaningful value to the value of interpreting the means in relation to the goal. In other words, the prior distribution describes the goal, the likelihood describes the means, and the posterior distribution describes interpretation. By taking a goal as a prior belief and observing an instrument as a means, the meaning of the instrument is interpreted as a posterior belief.

To systematically describe value generation across its full cycle, this study employs the free energy principle, which describes perception, learning, and action as the minimization of free energy (Friston et al., 2006; Friston, 2010). By mapping free energy minimization onto the three processes of the Triadic Value Model, this work provides a unified basis for deriving practical design methodologies.

2.2 Free energy principle

The free energy principle explains perception, learning, and action as the optimization of parameters of an internal generative model that represents the external world, such that a quantity known as free energy—an index of prediction error—is minimized (Friston et al., 2006; Friston, 2010). Free energy can be expressed as follows (Yanagisawa, 2021; Yanagisawa & Honda, 2025).

$$\begin{aligned} F &= E_{q(s)}[\ln q(s) - \ln p(s,o)] \\ &= D_{KL}[q(s)\|p(s|o)] + D_{KL}[p(s|o)\|p(s)] + E_{p(s|o)}[-\ln p(o|s)] \end{aligned} \tag{2}$$

Here, $q(s)$ denotes the recognition distribution, and $p(s,o)$ denotes the generative model. Free energy is defined as the expectation, under the recognition distribution, of the difference between the logarithms of the recognition distribution and the generative model. The free energy principle explains perception as the optimization of the recognition distribution so as to minimize free energy. The first term corresponds to the Kullback–Leibler divergence between the recognition distribution and the posterior distribution. When the recognition distribution is optimized to match the posterior distribution, the Kullback–Leibler divergence becomes zero. In other words, the Kullback–Leibler divergence represents the component of free energy that is reduced through perception. Assuming that, prior to perception, the recognition distribution corresponds to the prior distribution, the free energy reduced by perception (KLD) can be expressed as follows.

$$KLD = D_{KL}[p(s)\|p(s|o)] \tag{3}$$

The second term is the Kullback–Leibler divergence of the posterior distribution with respect to the prior distribution, referred to as Bayesian surprise (BS).

$$BS = D_{KL}[p(s|o)\|p(s)] \tag{4}$$

When the prior distribution is updated to the posterior distribution, BS becomes zero. BS represents the component of free energy that is reduced through learning (Bayesian updating). Bayesian updating is a statistical procedure in which the posterior distribution is used as the prior distribution for the next perceptual step and is updated to a new posterior distribution by observing new data.

The third term represents observational uncertainty (U), defined as the expected negative log-likelihood under the posterior distribution. While U can be reduced through action, future observations are unknown prior to action. Accordingly, action is described in terms of minimizing expected free energy, as formalized in active inference (Friston et al., 2016; Parr et al., 2022).

2.3 Active inference

Active inference explains action as the minimization of expected free energy (Friston et al., 2016; Parr et al., 2022). Expected free energy can be expressed as follows.

$$\begin{aligned} G &= E_{p(o|s)}E_{q(s|\pi)}[\ln q(s|\pi) - \ln p(s,o)] \\ &= D_{KL}[q(s|\pi)\|p(s)] - E_{q(o|\pi)}\big[D_{KL}[q(s|o,\pi)\|q(s|\pi)]\big] + E_{q(o|\pi)}[-\ln q(o|\pi)] \end{aligned} \tag{5}$$

Expected free energy is the expectation of free energy with respect to the likelihood, which is a function of observations. Here, π denotes an action. The recognition distribution is the probability distribution over states conditioned on action. Expected free energy can be decomposed into three terms. The first term is the Kullback–Leibler divergence between the recognition distribution and the prior distribution, which represents the discrepancy between the prior distribution predicted under a given action and the actual prior distribution (Risk). The second term is the expected value, with respect to the predicted observation distribution under a given action, of Bayesian surprise (Epistemic Value). The third term is the entropy of the observation probability distribution predicted under a given action.

Minimization of expected free energy is achieved through minimization of Risk, maximization of Epistemic Value, and minimization of entropy. Risk minimization is achieved when the prior distribution predicted under a given action matches the actual prior distribution. In other words, Risk is minimized by selecting actions that fulfill prior beliefs, which corresponds to exploitative behavior. In contrast, maximization of Epistemic Value is achieved by maximizing the Bayesian surprise predicted under a given action and therefore corresponds to exploratory behavior. Entropy minimization is achieved by selecting actions that minimize the uncertainty of predicted observations, corresponding to actions that reduce uncertainty in anticipated outcomes.

2.4 Correspondence between information measures and TVM

The processes described by the free energy principle and active inference are mapped onto the processes of TVM. Specifically, information-theoretic quantities in the free energy principle and active inference are associated with the value-generation processes in TVM.

KLD represents the component of free energy that is reduced through perception. Because it is reduced in the process of generating a posterior distribution from observations, KLD is considered to correspond to the process of generating meaningful value by assigning meaning to instrumental value.

BS represents the component of free energy that is reduced through learning (Bayesian updating). Because it is reduced in the process of updating the prior distribution through the posterior distribution, BS is considered to correspond to the process of updating intrinsic value by legitimizing meaningful value.

Expected free energy is the expectation of free energy with respect to the likelihood, which is a function of observations, and is minimized through action. Because it represents the information quantity minimized in the process of generating observations through action based on a prior distribution, expected free energy is considered to correspond to the process of generating instrumental value through the embodiment of intrinsic value. Minimization of expected free energy is achieved through minimization of Risk, maximization of Epistemic Value, and minimization of entropy. This implies that, in the process of generating instrumental value through the embodiment of intrinsic value, there exist exploitative actions that generate instrumental value to fulfill intrinsic value, exploratory actions that maximize the updating of intrinsic value through the legitimization of meaningful value, and actions that minimize the uncertainty of instrumental value.

2.5 Gaussian approximation

In order to construct practical methodologies corresponding to each process of TVM, the prior distribution, likelihood, and action-conditioned recognition distribution required for computing each information measure are represented using explicit probability distributions. Specifically, Gaussian distributions are adopted.

When the prior distribution, likelihood, and action-conditioned recognition distribution are approximated by Gaussian distributions, they can be expressed as follows.

$$p(s) = N\left(\mu_p, \sigma_p^2\right) \tag{6}$$

$$p(o|s) = N\left(s, \sigma_l^2\right) \tag{7}$$

$$q(s|\pi) = N(\mu_\pi, \sigma_\pi^2) \tag{8}$$

Here, μ_p denotes the mean of the prior distribution, σ_p^2 denotes the variance of the prior distribution, σ_l^2 denotes the variance of the likelihood, μ_π denotes the mean of the action-conditioned recognition distribution, and σ_π^2 denotes the variance of the action-conditioned recognition distribution. By approximating the prior distribution, likelihood, and action-conditioned recognition distribution with Gaussian distributions, the information measures corresponding to each process of TVM can be computed.

3 RESULT

By substituting Eqs. (6) and (7) into Eq. (3), KLD can be expressed as follows.

$$KLD = \frac{1}{2}\left(\frac{\sigma_p^2\left(o-\mu_p\right)^2}{\sigma_l^2\left(\sigma_p^2+\sigma_l^2\right)} - \ln\frac{\sigma_p^2+\sigma_l^2}{\sigma_l^2} + \frac{\sigma_p^2}{\sigma_l^2}\right) \tag{9}$$

By substituting Eqs. (6) and (7) into Eq. (4), BS can be expressed as follows:

$$BS = \frac{1}{2}\left(\frac{\sigma_p^2\left(o-\mu_p\right)^2}{\left(\sigma_p^2+\sigma_l^2\right)^2} + \ln\frac{\sigma_p^2+\sigma_l^2}{\sigma_l^2} - \frac{\sigma_p^2}{\sigma_p^2+\sigma_l^2}\right) \tag{10}$$

Figure 1 shows how the KLD and BS vary as functions of the variance of the prior distribution (σ_p^2) and the variance of the likelihood (σ_l^2). Across all conditions, KLD increases monotonically with increasing σ_p^2. In contrast, increasing σ_l^2 consistently reduces KLD. BS exhibits a similar dependence on σ_l^2, decreasing as the likelihood variance increases. With respect to the variance of the prior distribution, BS varies more gradually than KLD and shows a weaker sensitivity to σ_p^2 within the range visualized in Figure 1. Taken together, Figure 1 indicates that the maximization of both KLD and BS is facilitated by increasing the uncertainty of prior beliefs and improving the precision of observations.

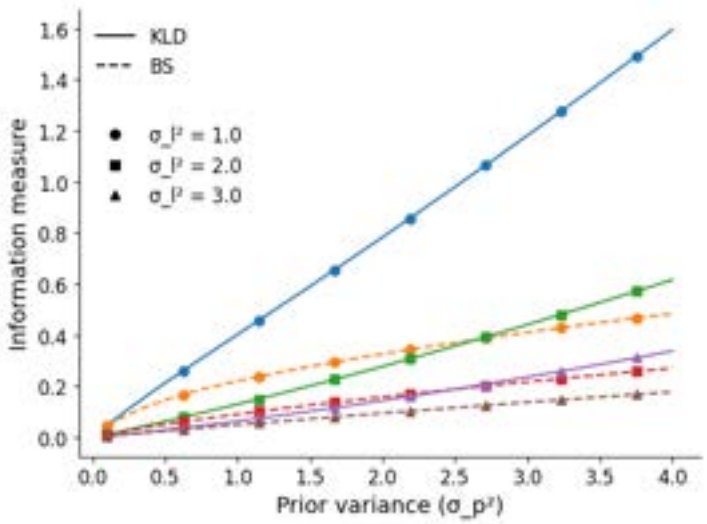


Figure 1. Effects of prior variance (σ_p^2) and likelihood variance (σ_l^2) on Kullback–Leibler divergence (KLD; solid lines) and Bayesian surprise (BS; dashed lines). Marker shapes indicate different values of σ_l^2.

By substituting Eqs. (6), (7), and (8) into Eq. (5), expected free energy can be expressed as follows:

$$G = \frac{1}{2}\left(\frac{\left(\mu_\pi - \mu_p\right)^2}{\sigma_p^2} + \ln\frac{2\pi\sigma_p^2\sigma_l^2}{\sigma_\pi^2} + \frac{\sigma_\pi^2}{\sigma_p^2}\right) \tag{11}$$

Figure 2 illustrates the relationship between expected free energy (G) and the variance of the action-conditioned belief (σ_π^2), with the prior variance held constant. The curve exhibits a clear U-shaped profile, with a minimum occurring when σ_π^2 is equal to σ_p^2. When σ_π^2 is smaller than σ_p^2, expected free energy increases. Conversely, when σ_π^2 exceeds σ_p^2, expected free energy also increases. These results indicate that expected free energy is minimized when the uncertainty of beliefs about outcomes predicted under a given action is comparable to the uncertainty of prior beliefs.

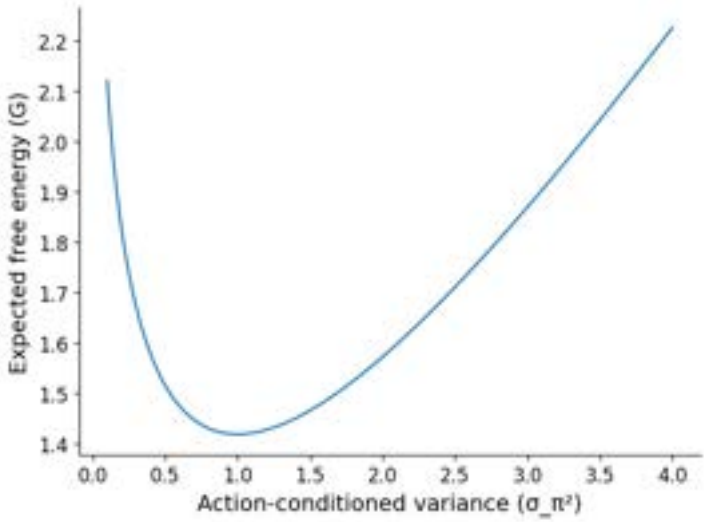


Figure 2. Expected free energy (G) as a function of action-conditioned variance (σ_π^2), showing a minimum when σ_π^2 is comparable to σ_p^2.

4 DISCUSSION

The variance of the prior distribution (σ_p^2) reflects the openness or ambiguity of intrinsic goals. The variance of the likelihood (σ_l^2) corresponds to the clarity or reliability of instrumental value embodied in artifacts. The variance of the action-conditioned belief (σ_π^2) represents the uncertainty of outcomes predicted to result from a given action. In design practice, this corresponds to how widely the consequences of an intervention—such as a prototype or scenario—are expected to vary, rather than the action itself. Framing innovation of meaning as the coordinated manipulation of these three variances allows designers to systematically balance inquiry and realization.

The process of innovation of meaning consists of three stages: identifying new meaningful value from existing instrumental value, updating existing intrinsic value through newly generated meaningful value, and optimizing existing instrumental value according to newly updated intrinsic value. For example, in the case of a candle, a designer identified a new meaning of self-care from the traditional candle, updated the goal from lighting to relaxation based on this meaning, and optimized the candle into an aromatherapy candle to serve the new goal. In the Method section, the process from instrumental value to meaningful value was associated with KLD, the process from meaningful value to intrinsic value with BS, and the process from intrinsic value to instrumental value with expected free energy. Accordingly, the process of innovation of meaning is interpreted as the maximization of KLD, the maximization of BS, and the minimization of expected free energy.

The maximization of KLD and BS corresponds to the epistemic emotions that drive the inquiry cycle proposed by Yanagisawa and Honda (2025). The minimization of expected free energy represents a balance between exploratory and exploitative behavior. Exploratory behavior corresponds to actions that maximize predicted epistemic emotions, whereas exploitative behavior corresponds to the generation of instrumental value that fulfills a given goal. In this sense, the minimization of expected free energy entails both the activation of the inquiry cycle and the achievement of goals.

As shown in Figure 1, the combined magnitude of KLD and BS increases as the variance of the prior distribution (σ_p^2) increases and decreases as the variance of the likelihood (σ_l^2) increases. This visual trend is consistent across the parameter ranges examined and indicates that the joint processes of meaning assignment and value updating are amplified when prior beliefs are less constrained while observational precision is high. From this perspective, Figure 1 suggests that innovation of meaning is facilitated by loosening commitment to existing goals and simultaneously attending to reliable observational cues, allowing newly generated meanings to exert a stronger influence on the updating of intrinsic value.

Figure 2 indicates that expected free energy is minimized when the variance of predicted outcomes conditioned on action (σ_π^2) is comparable to the variance of prior beliefs (σ_p^2). This condition reflects a balance between exploratory and exploitative actions in design. When σ_π^2 is much smaller than σ_p^2, designers favor highly predictable outcomes, resulting in conservative actions that limit exploration. In contrast, when σ_π^2 is much larger than σ_p^2, actions prioritize exploration at the expense of reliably realizing intrinsic value. The condition $\sigma_\pi^2 \approx \sigma_p^2$ therefore represents an optimal balance, in which designers explore new possibilities while still effectively exploiting existing beliefs to generate instrumental value in innovation of meaning.

In summary, the maximization of KLD and BS and the minimization of expected free energy are mapped onto the process of innovation of meaning. The maximization of KLD and BS is driven by weakening prior beliefs and increasing observational precision. The minimization of expected free energy is achieved by generating instrumental value through actions that are predicted to produce outcome variances comparable to those of prior beliefs. These design-relevant guidelines are expected to inform an innovation of meaning process that optimally balances exploration and goal achievement.

The present analysis relies on a Gaussian (Laplace) approximation of distributions, which is valid only in the vicinity of the mode. Consequently, the derived design guidelines characterize value-generation dynamics primarily around expected meanings and do not fully capture tail phenomena such as radical reinterpretations. As discussed by Yanagisawa and Honda (2025), such tail behaviors require the introduction of non-Gaussian components, such as uniform distributions. Therefore, the results of this study should be interpreted as local, expectation-centered guidelines rather than global guarantees for all innovation of meaning processes.

5 CONCLUSION

In this study, the three value-generation processes described by the Triadic Value Model (TVM) were formalized based on the free energy principle and active inference, and were reconstructed as a practical methodology. Specifically, the process of assigning meaning to instrumental value was associated with Kullback–Leibler divergence, the process of updating intrinsic value through the legitimization of meaningful value with Bayesian surprise, and the process of generating instrumental value through the embodiment of intrinsic value with expected free energy.

Through an analysis based on Gaussian approximation, this study clarified how each information measure varies with the variances of the prior distribution, the likelihood, and the action-conditioned recognition distribution. The results demonstrate that innovation of meaning can be understood as the maximization of KLD and BS and the minimization of expected free energy. The maximization of KLD and BS drives the inquiry cycle by weakening prior beliefs while increasing observational precision. In contrast, the minimization of expected free energy provides action guidelines that appropriately balance exploratory and exploitative behaviors, thereby enabling the simultaneous achievement of goals and updating of values.

ACKNOWLEDGMENTS

This research was conducted with support from the Social Collaboration Program between Mitsubishi Electric Corporation and the University of Tokyo. We extend our gratitude here.

REFERENCES

Yanagisawa, H. (2026). Reframing value as inference: The triadic value model for meaning-centered design, International Journal of Affective Engineering (in press).

Kushi, S., Nakaji, K., & Yanagisawa, H. (2024). MEANING GENERATION THEORY: A BAYESIAN APPROACH TO SIGN× CONTEXT= MEANING. In DS 136: Proceedings of the Asia Design and Innovation Conference (ADIC) 2024 (pp. 019-027).

Verganti, R. (2009). Design driven innovation: changing the rules of competition by radically innovating what things mean. Harvard Business Press.

Norman, D. A., & Verganti, R. (2014). Incremental and radical innovation: Design research vs. technology and meaning change. Design issues, 30(1), 78-96.

Kushi, S., & Yanagisawa, H. (2024). Innovation of meaning: design-driven study based on the interpretive theory of new meaning. Proceedings of the Design Society, 4, 35-44.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences, 27(12), 712-719.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. Journal of physiology-Paris, 100(1-3), 70-87.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

Yanagisawa, H. (2021). Free-energy model of emotion potential: Modeling arousal potential as information content induced by complexity and novelty. Frontiers in computational neuroscience, 15, 698252.

Yanagisawa, H., & Honda, S. (2025). Modeling the arousal potential of epistemic emotions using Bayesian information gain: a framework for inquiry cycles driven by free energy fluctuations. Frontiers in Psychology, 16, 1438080.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. Neuroscience & Biobehavioral Reviews, 68, 862-879.

Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active inference: the free energy principle in mind, brain, and behavior. MIT Press.

KEER2026, EUGENE, OR, USA| AUGUST 11-12, 2026
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2026

A MULTI-LAYER PSYCHOPHYSICAL MODEL OF PREFERENCE FOR FREQUENCY RESPONSE (EQ) IN MUSIC LISTENING

Taiki SHIMOZONO[1a], Yasuo SHIOZAWA[a] , Yuma YANO[b], Kiichi NAKA[c], Yoichi YAMAZAKI[d] , Noriko NAGATA[b]

[a] *Yamaha Corporation, Japan, taiki.shimozono@music.yamaha.com*
[b] *Kwansei Gakuin University, Japan, y-yuma1518@kwansei.ac.jp, nagata@kwansei.ac.jp,*
[c] *University of Nagasaki, Japan, naka.kiichi.research@gmail.com*
[d] *Aichi Prefectural University, Japan, yamazaki@ist.aichi-pu.ac.jp*

ABSTRACT

Frequency response (EQ) characteristics strongly affect listening satisfaction in music playback, but why a given EQ is preferred may differ depending on a listener's evaluation perspective. This study models the EQ preference process as a multi-layer psychophysical model—acoustic features → impressions → value/emotion—using both qualitative and quantitative approaches. First, we extracted individual evaluation structures for "preference" using the Evaluation Grid Method (EGM) and screened the resulting terms into a set of 43 rating words (16 value/emotion terms and 27 impression terms). Factor analyses organized impressions into four factors (Power, Relaxation, Clarity, and Coherence) and value/emotion into three factors (Comfortable, Overwhelmed, and Brightened). We also derived principal components from acoustic features and estimated structural paths linking them to impressions and value/emotion, yielding a model with good fit and high explanatory power. Listeners were further classified into three user types—Emotion, Sound, and Environment—and the results suggest that pathways of value formation may differ across types.

Keywords: *EQ preference, multi-layer model, impression, affect/value, user segmentation*

[1] Corresponding author. West Tower, Yokohama Symphostage 5-1-2 Minatomirai, Nishi-ku, Yokohama, Kanagawa, Japan. taiki.shimozono@music.yamaha.com

1 INTRODUCTION

Listening satisfaction in music is influenced not only by musical content but also by the frequency response characteristics of playback systems, such as the frequency response of loudspeakers/headphones and equalization (EQ) adjustments. EQ is a common means of shaping perceived balance to achieve experiences described as “easy to listen to,” “powerful,” or “calming.” However, the relationship between EQ-induced acoustic changes and subjective preference or value judgments is not straightforward. Even under the same EQ manipulation, listeners may prioritize different evaluation perspectives, leading to divergent preference judgments.

Previous studies have often addressed EQ and preference statistically, focusing on which responses tend to be preferred. In contrast, less attention has been paid to why they are preferred—namely, the evaluative process through which acoustic changes are translated into impressions and ultimately into value or emotion. To address this gap, the present paper (1) extracts evaluation structures of preference using the Evaluation Grid Method (EGM), (2) constructs a multi-layer psychophysical model based on rating experiments and acoustic-feature analysis, and (3) compares the model across user types to clarify individual differences and their underlying evaluative pathways.

2 RELATED WORK

2.1 Frequency Response and Preference

Equalization (EQ) is one of the primary factors shaping the perceived sound quality of headphones and loudspeakers. Prior work has reported both a commonly preferred tendency (often described as a target response) and substantial individual differences around that tendency. For example, Olive et al. (2014) conducted double-blind listening tests across multiple countries and demonstrated how frequency-response tuning influences sound-quality ratings and preference, as well as how preference varies with listeners’ age, experience, and cultural background. In addition, Olive (2019) applied clustering to preference data and showed that listeners can be segmented based on their preferred frequency-response profiles, with different segments exhibiting distinct response-shape tendencies (e.g., emphasis in low- and/or high-frequency regions).

2.2 Measuring Impressions and Emotion in Audio/Music (SD and Affect Grid)

Subjective evaluation of sound and music has been studied using two complementary frameworks: representing verbal impressions with a small number of underlying dimensions, and describing affective states in a low-dimensional space. A representative approach for the former is the Semantic Differential (SD) method, whose major dimensions—evaluation, potency, and activity—are widely recognized as a basic interpretive reference for impression ratings (Katahira et al., 2016). For the latter, affect is often characterized along two dimensions, valence and arousal, positioned as core components of emotional experience (core affect) (Ono et al., 2019). In this study, we use both perspectives—organizing impressions with reference to evaluation/potency/activity and representing affect in terms of valence and arousal—as a conceptual basis for describing the evaluative process from EQ-induced changes to preference judgments.

2.3 Uses of Music and Individual Differences

Music listening is not only a matter of "like/dislike"; its meaning also depends on *how* and *why* people listen, such as the intended purpose and situation of listening. Chamorro-Premuzic and Furnham (2007) investigated everyday functions of music listening and showed that uses of music can be organized into dimensions such as emotional use, cognitive use, and background use. This perspective suggests that preference for EQ may likewise reflect not only acoustic differences per se but also evaluation viewpoints tied to listeners' situations and listening purposes. Accordingly, we position the "uses of music" perspective as an interpretive framework for understanding evaluation structures and for defining user types in the present study.

3 RESEARCH FRAMEWORK: DEFINITION OF THE MULTI-LAYER PSYCHOPHYSICAL MODEL

Building on prior work showing that preferred frequency-response profiles vary across listeners, and drawing on frameworks for impression and affective evaluation, this study focuses on the internal evaluative pathways through which acoustic changes are interpreted as impressions and value/emotional responses. The proposed three-layer model consists of (i) an **acoustic feature layer** representing signal properties altered by EQ, (ii) an **impression layer** representing listeners' subjective impressions, and (iii) a **value layer** representing value judgments and affective responses elicited during listening. The impression layer refers to the major dimensions discussed in the Semantic Differential framework (evaluation, potency, and activity), and the value layer refers to the two-dimensional core-affect space (valence and arousal). We primarily assume a mediated pathway in which acoustic features influence value judgments through impressions, while not excluding the possibility of a direct influence from the acoustic feature layer to the value layer. Moreover, the mapping between layers—i.e., how physical differences are interpreted as particular impressions and values—may depend on personal constructs formed through experience (Kelly, 1955). Accordingly, we assume that the pathways between layers exhibit individual differences, as illustrated in the Figure1. A similar hierarchical approach to structuring impression layers has been demonstrated in perceptual evaluations across domains, including music liveness (Shimozono et al., 2025) and 3D shapes (Miyai et al., 2019).

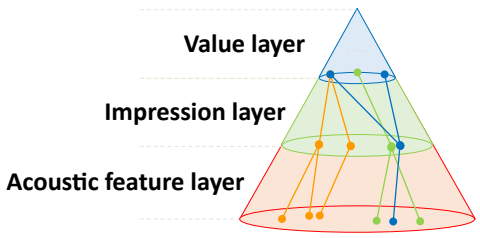


Figure 1. The multi-layer psychophysical model assumed in this study

4 EVALUATION STRUCTURE ANALYSIS

4.1 Aim

This chapter extracts individual evaluation structures for EQ preference through semi-structured interviews using the Evaluation Grid Method (EGM).

4.2 Stimuli

4.2.1 Source music

We selected the musical pieces from the Music Genre Database of the RWC Music Database (RWC-MDB-G), a copyright-cleared database designed for shared use in research (Goto et al., 2003) (Table 1).

Table 1. Audio stimuli used

Gener	ID	Title	Composer	SPL(dBA)
Classic	RWC-MDB-G-051	Prelude to Act I from "Die Meistersinger von Nürnberg"	Wagner, R.	70.4
Dance	RWC-MDB-G-021	Stf	Yamamoto, Ichi	67.6
Jazz	RWC-MDB-G-033	The Ramp	Haji, Hiroshi	66.6
Pop	RWC-MDB-G-003	In Your Arms	Kawagoe, Yu	75.3
Rock	RWC-MDB-G-008	Living in a World of Destruction	Kamaki, Shigeka	74.1

4.2.2 EQ processing

We prepared six EQ conditions: Flat (no EQ), low-frequency boost, low-frequency cut, mid-frequency boost, mid-frequency cut, and high-frequency cut. The low-frequency conditions were implemented as a low-shelf filter at 250 Hz (Q/slope = 1.0) with a gain of ±9 dB. The mid-frequency conditions were implemented as a peaking filter at 1 kHz (Q = 0.5) with a gain of ±9 dB. The high-frequency condition was implemented as a high-shelf filter at 4 kHz (Q/slope = 1.0) with a gain of −9 dB (Figure 2).

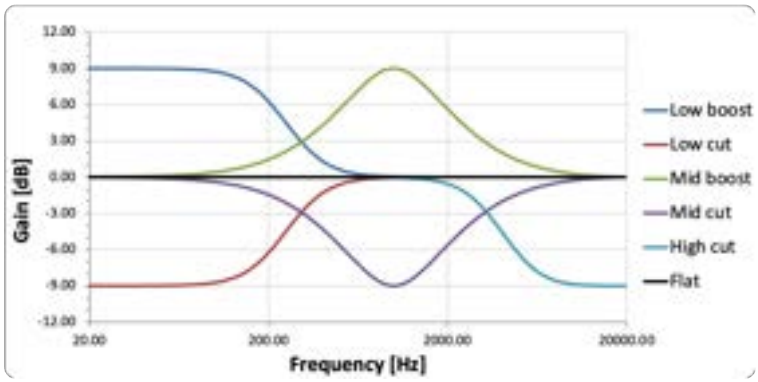


Figure 2. Frequency response curves of the six EQ conditions

4.2.3 Loudness normalization

Because level differences can strongly influence preference judgments in subjective evaluation, stimulus levels must be controlled. Stimuli were normalized using loudness estimation based on ISO 532-1 (Zwicker method) to minimize inter-stimulus level differences (ISO, 2017).

4.3 Experimental Design

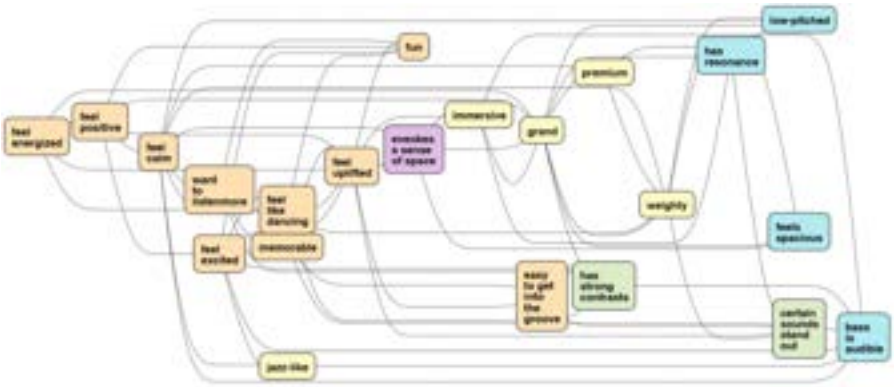
Participants were 20 university and graduate students (11 male, 9 female; age range = 19–23 years, M = 20.9, SD = 1.0) with normal hearing. All participants were right-handed. Their musical backgrounds varied: 7 were classified as musicians and 13 as non-musicians, and their daily music listening time ranged from 0.5 to 5.0 h (M = 2.77, SD = 1.18). We prepared 30 stimuli in total (5 genres × 6 EQ conditions). The experiment was conducted in a quiet environment, with

participants seated and listening via headphones. The presentation level was set to approximately 70 dBA. The study was exploratory, aiming to elicit evaluation viewpoints under controlled listening conditions and examine the feasibility of the proposed multi-layer framework. The sample size should therefore be considered a limitation when generalizing the user types and path models.

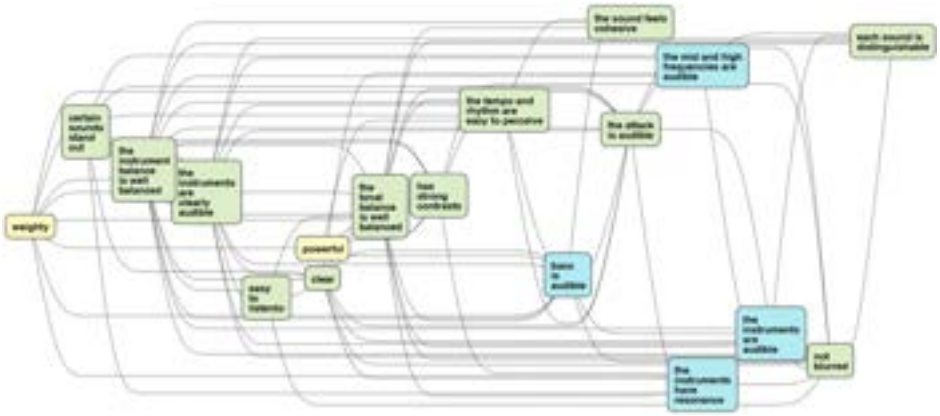
Within EGM, comparisons were made within each genre, i.e., participants compared only the six EQ conditions of the same genre (e.g., the six “Classic” stimuli were compared with each other). For each genre, participants listened to the six conditions and first selected the condition they found most preferable. They then assigned that condition a preference score of 100, and rated the remaining five conditions on a 0–100 scale relative to it. After the ratings, we conducted EGM interviews focusing on preference within the same genre. The interviews used laddering (ladder-up/ladder-down) to explore higher-level concepts and lower-level concrete reasons starting from participants’ own expressions (Sanui, 1996; Reynolds & Gutman, 1988; Sugimoto et al., 2023).

4.4 Results

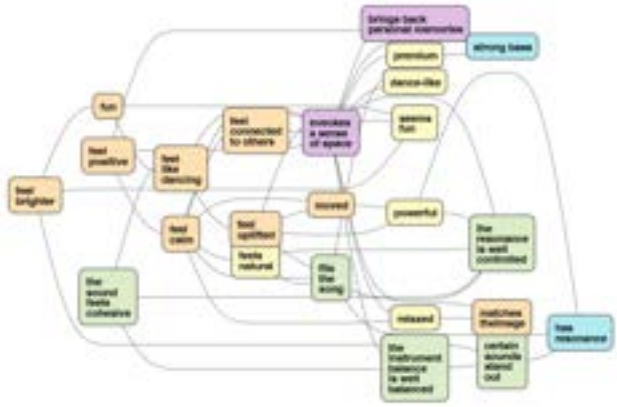
From the EGM interview records, we extracted utterances related to preference judgments as evaluation terms, yielding a total of 3,756 terms. Rather than directly referring to frequency bands or EQ operations, many terms described perceptual attributes and musical elements (e.g., how instruments are heard, contours, texture, and balance). Based on the evaluation-term data and individual evaluation structures, we found that listeners relied on different dominant viewpoints even under the same EQ conditions. To quantify these differences, we modeled each participant’s evaluation structure as a network and used Katz centrality to extract central terms (Katz, 1953). We then assigned each participant to a dominant viewpoint type, defined as the viewpoint category (emotion, sound, or environment) that accounted for the largest share of the extracted central terms. Consequently, participants were classified into 6 Emotion-type, 11 Sound-type, and 3 Environment-type listeners; the corresponding evaluation-structure diagrams are shown in Figure 3. These types may be interpreted in relation to prior frameworks on uses/functions of music listening (Chamorro-Premuzic & Furnham, 2007).



(a) Emotion type



(b) Sound type



(c) Environment type

Figure 3. Representative evaluation-structure networks for the three user types

5 QUALITATIVE MODEL DEVELOPMENT

Because the 3,756 EGM-derived evaluation terms were too numerous for subsequent rating experiments and modeling, this chapter constructs a representative rating-term set through (1) extracting central terms based on Katz centrality (Katz, 1953), (2) screening terms for usability in expressing EQ-induced differences, and (3) grouping and merging semantically similar terms. The final set comprised 43 terms (16 value/emotion terms and 27 impression terms).

Table 2. Selected evaluation terms for subsequent rating experiments

Value	Arousal	Comfort	Interest	Positivity	Moved	Naturalness/Ease	
preferred	uplifted	comfortable	memorable	brightened	cool	natural	
good	excited	pleasant	has substance	energized	overwhelmed	easy to listen to	
satisfied		reassured					
Clarity	**Presence**	**Audibility**	**Spatiality**	**Stability**	**Power**	**Coherence**	**Warmth**
clear	feels exciting	instruments clearly audible	has depth	stable	weighty	well-balanced	warm
each sound is distinguishable	immersive	bass is audible	has resonance		powerful	cohesive	
Imagery	**Ease of listening**	**Contrast**	**Sustain**	**Familiarity**	**Tempo /rhythm**	**Luxury**	
fits the song	easy to hear	has strong contrasts	good sustain	approachable	tempo and rhythm are perceivable	luxurious	
evokes space	good sound quality	penetrating		familiar	relaxed	premium	

6 QUANTITATIVE MODEL DEVELOPMENT

6.1 Aim

This chapter estimates a multi-layer structural model linking acoustic features to impressions and value/emotion responses.

6.2 Rating Experiment for Value/Emotion and Impressions

To quantify value/emotion and impressions, we conducted a headphone-based rating experiment. The stimuli were the same 30 audio excerpts as in Section 4.3 (5 genres × 6 EQ conditions). For each stimulus, participants rated both value/emotion and impressions using 43 evaluation terms identified in the previous chapter (16 value/emotion terms and 27 impression terms). Ratings were made on a 7-point scale for each stimulus–term combination.

6.3 Factor Analysis of Impression Structure and Value/Emotion Structure

6.3.1 Impression Structure (27 impression terms)

We examined several criteria to determine the number of factors and selected a solution based on interpretability. Factors were extracted using maximum likelihood estimation and rotated with Promax rotation. The impression ratings were summarized into four factors: Power (e.g., “powerful,” “weighty,” “immersive”), Relaxation (e.g., “relaxed,” “warm,” “friendly”), Clarity (e.g., “clear,” “each sound is discernible”), and Coherence (e.g., “fits the music,” “cohesive,” “well-balanced”).

6.3.2 Value/Emotion Structure (16 value/emotion terms)

The value/emotion ratings were analyzed in the same manner (maximum likelihood extraction with Promax rotation). The results were summarized into three factors: Comfortable (positive valence / low arousal), Overwhelmed (high arousal), and Brightened (positive valence / high arousal). The results are summarized in Table 3.

Table 3. Factor loadings for impression (27 terms; four factors) and value/emotion ratings (16 terms; three factors) (maximum likelihood, Promax rotation).

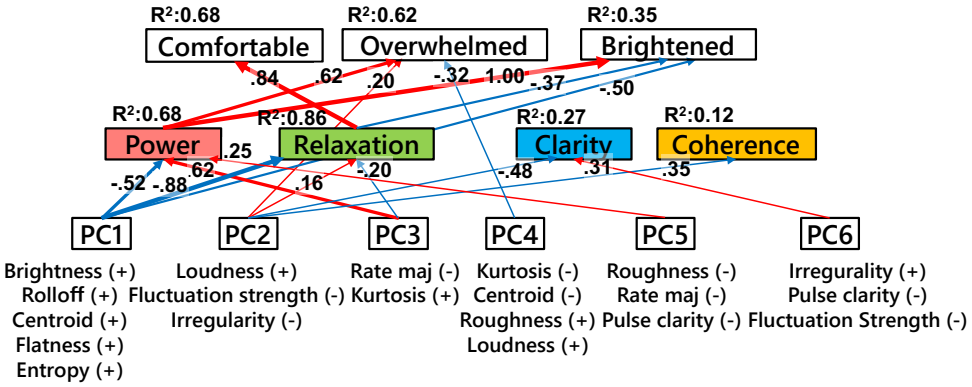


6.4 Acoustic Feature Extraction

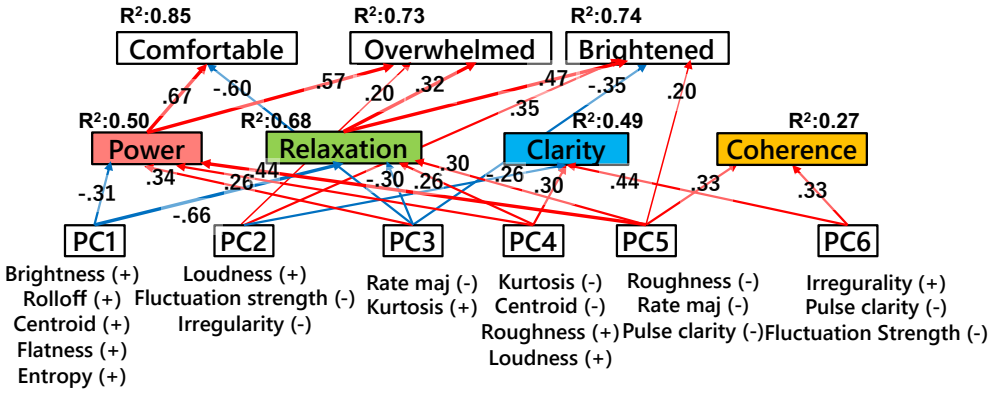
As explanatory variables capturing physical changes underlying impressions and values, we computed acoustic features from each stimulus using the MIR toolbox in MATLAB. We calculated 15 commonly used descriptors, mainly related to spectral information (e.g., loudness, sharpness, roughness, fluctuation strength; spectral statistics such as spectral centroid and flatness; pulse clarity; and major ratio). Because these descriptors are mutually correlated, we applied principal component analysis for dimensionality reduction. Based on cumulative explained variance, the first six principal components accounted for 99.5% of the variance; therefore, PC1–PC6 were used as representative acoustic variables in subsequent modeling.

7 USER-TYPE DIFFERENCES IN VALUE-FORMATION PATHWAYS

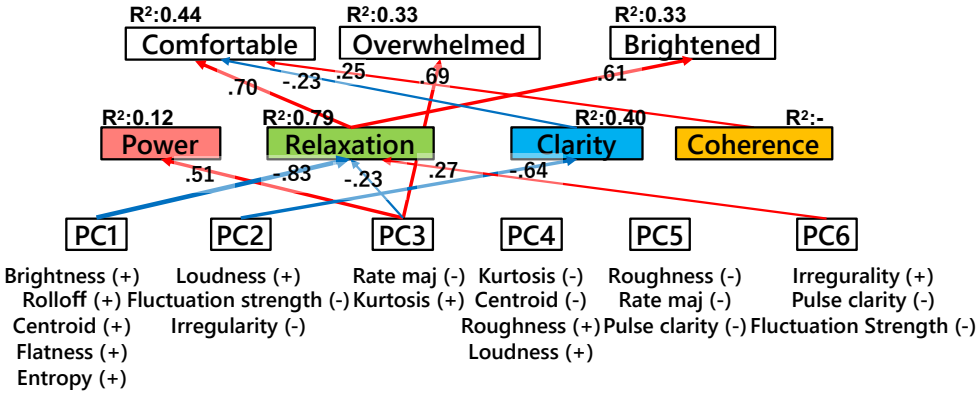
This chapter examines how value-formation pathways differ across the three user types (Emotion, Sound, and Environment) by estimating type-specific structural models. Figure 4 shows the value-structure models estimated for each user type. Model fit was generally good for all types: Emotion type (GFI = 0.97, AGFI = 0.92), Sound type (GFI = 0.97, AGFI = 0.88), and Environment type (GFI = 0.94, AGFI = 0.86).



(a) Emotion type



(b) Sound type



(c) Environment type

Figure 4. Structural models for each user type (emotion, sound, environment).

8 DISCUSSION

8.1 Significance of a Multi-layer Model for Representing EQ Preferences

In this study, EQ preference was modeled as a multi-layer process—acoustic features (principal components) → impressions → values—and we showed that the same modeling framework can be established for all three user types. This framework allows preference to be described not merely as a stronger or weaker liking, but as which impression cues are referenced and how they are translated into value responses (valence and arousal).

8.2 User-type-specific Approaches

Across types, the degree to which acoustic features explained impressions and values tended to be highest for the Sound type, followed by the Emotion type, and lowest for the Environment type. This implies that, particularly for the Environment type, contributions from factors that are difficult to capture by acoustic features alone—such as imagined situations and listening contexts—may be larger. In addition, the pathway forming the key value factor “Comfortable” differed by type: for the Sound type, the dominant pathway was mediated by Power (potency-related impressions), whereas for the Emotion and Environment types it was primarily mediated by Relaxation (low-activity impressions). From a practical standpoint, this suggests that EQ tuning strategies may be adapted by type: emphasizing a “power-oriented” direction for the Sound type, and a “relaxation-oriented” direction for the Emotion and Environment types. More broadly, the findings suggest that audio design can benefit from considering not only an average target response but also the subjective pathway through which users experience sound. If such evaluation viewpoints can be estimated or reflected in the design process, personalized or adaptive audio systems could support desired experiences—such as comfort, excitement, or ease of listening—through different acoustic strategies. Thus, the proposed model provides an interpretable basis for linking acoustic control parameters with user-perceived impressions and value responses.

9 CONCLUSION

This study showed that EQ preference can be modeled as a multi-layer process linking acoustic features, impressions, and value/emotion, with generally acceptable model fit. Categorizing participants into three user types (Emotion, Sound, and Environment) further revealed that value-formation pathways—such as the drivers of "Comfortable"—differ across types, supporting the usefulness of user-type-aware understanding and prediction. Future work should expand the sample size and experimental conditions, and integrate contextual factors such as use cases and listening environments.

REFERENCES

1. Olive, S. E., Welti, T., & McMullin, E. (2014). The influence of listeners' experience, age, and culture on headphone sound quality preferences. Audio Engineering Society Convention Paper.
2. Olive, S. E. (2019). Segmentation of listeners based on their preferred headphone sound quality profiles. Audio Engineering Society Convention Paper.
3. Katahira, K., Muto, K., Lee, N., Tobitani, K., Shiraiwa, A., Nakajima, K., Nagata, N., Kishino, F., Yamamoto, M., Kawasaki, K., Nikata, K., & Asano, T. (2016). Major factors in Kansei evaluation of 3D objects. *Transactions of Japan Society of Kansei Engineering*, *15*(4), 563–570. https://doi.org/10.5057/jjske.TJSKE-D-15-00087
4. Ono, Y., Aoyagi, S., Sugimoto, M., Yamazaki, Y., Yamamoto, M., & Nagata, N. (2019). Application of classification method of emotional expression type based on Laban movement analysis to design creation. In M. Kurosu (Ed.), *Human-Computer Interaction: Recognition and Interaction Technologies* (*Lecture Notes in Computer Science*, Vol. 11567, pp. 143–154). Springer. https://doi.org/10.1007/978-3-030-22643-5_11
5. Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology*, *98*(2), 175–185. https://doi.org/10.1348/000712606X111177
6. Kelly, G. A. (1955). *The psychology of personal constructs*. W. W. Norton.
7. Goto, M., Hashiguchi, H., Nishimura, T., & Oka, R. (2003). RWC music database: Music genre database and musical instrument sound database. *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*.
8. International Organization for Standardization. (2017). ISO 532-1:2017 Acoustics—Methods for calculating loudness—Part 1: Zwicker method.
9. Sanui, J. (1996). Visualization of users' requirements: Introduction of the Evaluation Grid Method. *Proceedings of the 3rd Design & Decision Support Systems in Architecture & Urban Planning Conference*, 365–374.
10. Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, *28*(1), 11–31. https://doi.org/10.1080/00218499.1988.12467766
11. Miyai, S., Katahira, K., Sugimoto, M., Nagata, N., Nikata, K., & Kawasaki, K. (2019). Hierarchical structuring of the impressions of 3D shapes targeting for art and non-art university students. In C. Stephanidis (Ed.), *HCI International 2019 – Posters Part I* (*Communications in Computer and Information Science*, Vol. 1032, pp. 385–393). Springer. https://doi.org/10.1007/978-3-030-23522-2_50
12. Shimozono, T., Sasai, D., Shiozawa, Y., Yamazaki, Y., Hashimoto, S., & Nagata, N. (2025). Understanding the structure of "liveness" and its value variations through large-scale affective evaluation. *Proceedings of the 18th International Conference on Music Perception and Cognition (ICMPC 18)*. Síntese Eventos.

Confidential

13. Sugimoto, M., Yagi, Y., & Nagata, N. (2023). How different tourist sites evoke different emotions: Investigation focusing on the urban and rural sites in Japan. In M. Kurosu & A. Hashizume (Eds.), *Human-Computer Interaction* (*Lecture Notes in Computer Science*, Vol. 14012). Springer. https://doi.org/10.1007/978-3-031-35599-8_21

14. Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43. https://doi.org/10.1007/BF02289026

AFFECTIVE DIFFERENCES IN DECISION PERFORMANCE BETWEEN EXPERIENCED AND INEXPERIENCED USERS IN FREE WALKING TOUR ONLINE BOOKING INTERFACES

TZU WEI SU[a], MING-YUAN MA[b]

[a] *National Cheng Kung University, Taiwan, mamy@mail.ncku.edu.tw*
[b] *National Cheng Kung University, Taiwan, bruce200102101002@gmail.com*

ABSTRACT

In the experience economy, evaluation standards have shifted from functional efficiency toward the capacity of services to evoke users' emotions and experiential value, thereby enhancing decision performance. However, despite tourism being a major experience-driven industry, most online booking interfaces continue to prioritize information density and operational efficiency, often overlooking differences in users' affective needs based on their familiarity with booking platforms.

Free Walking Tours (FWTs), a rapidly growing and unconventional tourism format, attract a substantial number of first-time participants. Inexperienced users frequently experience uncertainty and trust concerns, while experienced users focus on efficiently identifying tours that match their expectations. Function-oriented interfaces may therefore fail to support optimal decision performance for both groups.

This study applies Kansei Engineering, integrating morphological analysis and Quantification Theory Type I, to examine differences between experienced users (using booking platforms more than twice per year) and inexperienced users (less than once per year). The research aims to identify key interface design elements and affective attributes that enhance decision performance across user groups.

Findings indicate that inexperienced users prioritize friendliness and thoughtful consideration to reduce uncertainty, whereas experienced users emphasize security and reliability. Both groups value convenience and local authenticity. The results demonstrate that affectively aligned interface design improves decision quality, interaction fluency, and trust formation, providing strategic guidance for performance-oriented booking environments

Keywords: Kansei engineering , User Interface , Free Walking Tour , Tourist Experience , Affective Experience

1 INTRODUCTION

In the context of the experience economy, travelers increasingly seek tourism formats that offer deep engagement, authenticity, and local interaction. Free Walking Tours (FWTs), characterized by free pricing, multiple interpretive perspectives, and strong local connections,

have gradually emerged as one of the fastest-growing forms of urban tourism worldwide. Compared with standardized group tours, FWTs enable travelers to explore cities on foot and to understand local culture through guides' personal narratives and perspectives. Previous studies indicate that such walking tours can enhance tourists' learning, improve their appreciation of urban landscapes, and deepen overall travel experiences (Markwell, Stevenson, & Rowe, 2004).

1.1 The User Interface of Free Walking Tours

With the growth of youth independent travel and backpacker markets, the low cost, high flexibility, and strong social interaction of Free Walking Tours (FWTs) have made this format increasingly attractive. Urban walking tours have expanded rapidly over the past decade, as reflected in the growth of participation in cities such as Barcelona (Barcelona Tourism Bureau, 2014), indicating rising demand for FWTs as an innovative tourism product.

From an operational perspective, FWT services involve tour recruitment and tip-based compensation, with recruitment serving as the primary point of contact between travelers and the service. As tourism information seeking has become increasingly digitalized, FWT operators have shifted from street-based recruitment to online platforms, reviews, and social media. Accordingly, the FWT online interface has become a critical touchpoint for establishing trust, shaping first impressions, and influencing participation intentions.

1.2 Differences among Free Walking Tour Tourists

In the experience economy, emotional expectations increasingly influence users' decision performance in digital services. In the context of FWT online booking interfaces, user differences are primarily related to familiarity with travel reservation platforms rather than demographic factors.

In this study, Experienced users are defined as those who use travel booking websites more than twice per year, whereas Inexperienced users use them less than once per year. Differences in familiarity lead to distinct emotional needs that directly affect decision performance. Understanding how interface design evokes different emotional responses across these user groups is therefore essential for enhancing decision performance and participation intention.

Table 1. User Group Classification Criteria

Catergories	Frequency (a year)
Experienced users	>2
Inexperienced users	<1

1.3 Research Objectives

The objective of this study is to construct an emotionally aligned FWT online booking interface that enhances decision performance among users with different levels of platform familiarity. By clarifying the relationships among interface design elements, functional structures, and users' emotional preferences, this research aims to support the development of booking environments that respond to users' affective expectations rather than relying solely on functional efficiency.

Accordingly, this study addresses the following research questions:

- Q1: What interface functions and design categories are present in existing major FWT online booking platforms?
- Q2: Do experienced and inexperienced users exhibit differences in emotional perceptions toward FWT online booking interfaces?
- Q3: What key emotional attributes should FWT online booking platforms incorporate to enhance decision performance among experienced and inexperienced users?

2 LITERATURE REVIEW

2.1 The Three Levels of Design and Interface Design

Don Norman's Three Levels of Design provide a framework for understanding emotional responses in user–interface interaction, distinguishing among the visceral, behavioral, and reflective levels (Norman, 2004). In this study, results of morphological analysis are mapped onto these three levels to explain how interface characteristics influence tourists' emotional perceptions.

Specifically, visual elements and content presentation correspond to the visceral level, interaction behaviors such as page functions and navigation flows to the behavioral level, and emotional evaluations captured through Kansei words (e.g., convenience, friendliness, and immersion) to the reflective level. This alignment establishes the theoretical basis for applying Kansei Engineering and Quantification Theory Type I to link interface design features with tourists' emotional perceptions.

Table 2. Don Norman's Three Levels of Design

visceral level	behavioral level	reflective level
visuals, content, etc.	Interfaces usage flow	Affective responses
Influences users' first impressions and immediate intuitive responses to the interface.	Influences whether users can achieve their intended usage goals.	Influences whether users can achieve their intended usage goals.

2.2 Kansei Engineering Research Techniques

Kansei Engineering is a design methodology defined as "a technology for translating consumers' emotional needs and impressions into concrete design elements" (Mitsuo Nagamachi, 1995). This approach not only facilitates the systematic identification of human

emotional preferences but also supports their effective application in practical design contexts. The core concept of Kansei Engineering lies in collecting users' Kansei words and analyzing the relationships between these subjective emotional expressions and specific design features, thereby informing design decision-making.

Quantification Theory Type I is one of the most commonly used analytical methods within Kansei Engineering research. It is primarily employed to examine relationships between categorical independent variables and continuous dependent variables. Through this method, interface design morphological factors can be treated as categorical variables, allowing their relative influences on users' emotional perceptions to be quantitatively evaluated.

In the present study, Kansei Engineering combined with Quantification Theory Type I serves as a key methodological framework for linking tourists' subjective emotions with design elements of FWT online interfaces. This integrated approach enables a systematic analysis of emotional perception differences among diverse tourist groups and facilitates the translation of analytical results into concrete and actionable interface design strategies.

3 RESEARCH METHOD

This study integrates Kansei Engineering with morphological analysis to address the proposed research questions. The research procedure consists of four main stages: Case study & Sample Selection, Expert Interviews, Data collection, and Data analysis.

3.1 Case study & Sample Selection

Six representative Free Walking Tour (FWT) official websites and platform pages were selected as case studies from both domestic and international contexts. Selection was based on five criteria: High User Volume (HUV), High Brand Visibility (HBV), Multi-country Coverage (MCC), Interface Diversity (IDV), and Sustained Operation (SOP), ensuring representativeness and analytical relevance. Among the selected cases, Like It Formosa, the most representative FWT platform in Taiwan, was included to reflect the local context and enable cross-regional comparison.

Each platform was evaluated against the selection criteria using binary coding (0/1), where *1* indicates that the criterion is met and *0* indicates that it is not.

Table 3. Case Selection Results

	HUV	HBV	MCC	IDV	SOP
GuruWalk	1	1	1	1	1
SANDEMANs Tours	1	1	1	1	1
FREETOUR.com	1	1	1	1	1
Free Walking Tour	1	1	1	1	1
Like it formosa	1	1	1	1	1

3.2 Expert Interviews

Expert interviews were conducted with two Free Walking Tour (FWT) specialists and one user interface (UI) design expert. The interviews aimed to collaboratively identify key morphological

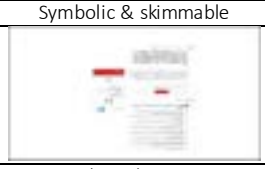
factors of FWT online marketing interfaces and to refine the selection of Kansei words used in the subsequent analysis. These interviews provided domain knowledge and design perspectives to ensure the relevance and validity of the analytical framework.

Table 4. Profiles of Interviewed Experts

	Field Expert	Years of Professional Experience
Mr. Cheng	Free Walking tourguide	5
Ms. Cheng	Free Walking tourguide	3
Ms. Zheng	UI Deisgner	4

3.2.1 Morphological Factor Analysis

Table 5. Morphological Factor Analysis

Interface usage flow	factor	level1	level2	Level3
CONTECT STAGE	TOUR SEARCH	Self-initiated search	Option-based browsing	Contextual inspiration
				
	CITY LIST	Visual-only	Single quantitative cue	Multiple quantitative cues
				
SELECT STAGE	TOUR LIST	Visual-only	Single quantitative cue	Multiple quantitative cues
				
INFORMATION SECTION	TOUR INFO	Textual narrative	Symbolic & skimmable	Structured overview
				
	TOUR MAP	Point-based reference	Stop-based structure	Area-based overview
				
	TOUR PHOTO	Supplementary	Visually driven	Parallel support
				
	TOUR REVIEWS	Low structured	Moderately structured	Highly structured
				

The FWT online marketing interface was analyzed using morphological factor analysis, distinguishing between interface forms and design factors. Interface forms represent shared functional sections and reflect the interface usage flow, while design factors correspond to UI design attributes within each section.

3.2.2 Kansei Word Collection

An initial set of eight Kansei words was generated to capture users' emotional perceptions of FWT online interfaces. Through KJ method clustering and expert discussions, these words were refined into a final set of six representative Kansei words for subsequent analysis.

3.2.3 Three-Level Correspondence Table

A Three-Level Correspondence Table was developed through expert discussions to map relationships between Kansei words and morphological factors based on Don Norman's three levels of design. This table provides a structured foundation for the subsequent Kansei Engineering analysis.

4 RESULT

The main thing to remember is: figures have the caption at the bottom, tables have the caption at the top.

4.1 Sample Characteristics and Kansei Evaluation

A total of 59 valid responses were collected. The sample is predominantly young, with most participants aged between 18–24, which aligns with typical Free Walking Tour users.

In terms of experience lvel, the sample consists of 40 experienced users (67.8%) and 19 first-time users (32.2%), enabling comparative analysis between user groups.

The Kansei evaluation results indicate generally positive perceptions across all dimensions. Among them, convenience (M = 4.21) received the highest score, while friendliness (M = 3.81) was the lowest.

This suggests that current FWT interfaces perform well in functional usability but are relatively weaker in emotional aspects, particularly in conveying friendliness and approachability.

Table 6. Demographic Characteristics of the Survey Sample

Category	Female	Male	18–24	25–34	35–44	45+	Experienced	First-time	Total
N	22	11	15	10	6	2	40	19	59
%	37.3	18.6	25.4	16.9	10.2	3.4	67.8	32.2	100

4.2 Quantification Theory Type I Analysis across User Segments

Quantification Theory Type I (QT1) was applied to examine the relationship between interface design factors and Kansei evaluations for both user groups.

For experienced users, the explanatory power is moderate (e.g., Safety $R^2 = 0.628$, Authenticity $R^2 = 0.539$), with results indicating that structure-related factors (e.g., x1, x2) have relatively stronger influence. This suggests that experienced users are primarily affected by efficiency and information organization.

In contrast, first-time users show consistently higher explanatory power across all Kansei dimensions ($R^2 = 0.618$–0.769), indicating that interface design plays a more significant role in shaping their perceptions. Multiple factors (x1–x4) exhibit strong effects, suggesting sensitivity to both structural and visual elements.

4.3 Comparison between User Groups

Comparative analysis of QT1 results reveals clear differences in how interface design factors influence emotional perception between experienced and first-time users.

For experienced users, the influence of interface factors is relatively moderate and concentrated on structural efficiency. Factors such as x1 (tour search) and x2 (tour list) show consistent but stable effects across Kansei dimensions. For example, structured browsing (x1–L2) enhances immersion, while clear list presentation (x2–L1) improves friendliness and safety. This indicates that experienced users rely on information organization and efficient navigation, and are less sensitive to variations in interface presentation.

In contrast, first-time users exhibit significantly stronger and more varied responses to interface design. Multiple factors (x1–x4) show large category scores across Kansei dimensions, indicating high sensitivity to specific design configurations. For example, option-based browsing (x1–L2) strongly increases friendliness but significantly reduces convenience, while contextual inspiration (x1–L3) enhances immersion and authenticity. Similarly, structured tour information (x4–L1) greatly improves thoughtfulness, whereas inappropriate configurations can lead to strong negative perceptions.

These results suggest that first-time users are influenced not only by structure, but also by how information is presented, contextualized, and guided. The presence of extreme positive and negative category scores further indicates that their experience is highly dependent on interface design quality.

Overall, the differences between the two groups highlight that experienced users prioritize efficiency and stability, whereas first-time users require both cognitive clarity and emotional support. This reinforces that the gap between user groups lies not only in usability, but in design sensitivity and emotional accessibility.

Table 7. Category Scores for Experienced and First-time Users

EXPEREINCED USER (n=40)							
Factor	Level	Friendliness	Convenience	Thoughtfulness	Immersion	Authenticity	Safety
x1	L1	0.188	0.013	-0.161	-0.099	0.279	-0.316
	L2	-0.203	0.004	0.093	0.505	-0.140	-0.082
	L3	0.028	-0.026	0.114	0.057	-0.131	0.277
x2	L1	0.688	0.378	-0.019	0.013	-0.153	0.395
	L2	-0.121	-0.142	0.000	0.157	0.071	0.128
	L3	0.081	0.215	0.007	-0.199	0.254	-0.342
x3	L1	—	-0.353	0.038	0.027	-0.198	-0.335
	L2	—	-0.314	0.031	-0.156	0.193	0.054
	L3	—	0.163	-0.036	0.211	0.111	0.133
x4	L1	—	—	0.238	-0.118	0.049	-0.043
	L2	—	—	-0.105	-0.086	0.015	0.122
	L3	—	—	-0.006	0.270	-0.029	-0.074
x5	L1	—	—	—	—	-0.080	0.082
	L2	—	—	—	—	-0.005	0.138
	L3	—	—	—	—	0.085	-0.257
x6	L1	—	—	—	—	—	-0.705
	L2	—	—	—	—	—	0.256
	L3	—	—	—	—	—	-0.100
x7	L1	—	—	—	—	—	-0.068
	L2	—	—	—	—	—	0.053
	L3	—	—	—	—	—	0.016
First-time Users (n = 19)							
Factor	Level	Friendliness	Convenience	Thoughtfulness	Immersion	Authenticity	Safety
x1	L1	0.035	0.204	0.479	-0.104	-0.402	—
	L2	0.624	-1.349	-0.793	—	-0.563	—
	L3	-0.733	-0.026	0.023	0.555	0.500	—
x2	L1	-2.255	-0.541	0.498	0.160	0.593	—
	L2	-0.043	0.066	-0.178	0.506	0.596	—
	L3	0.462	-0.015	0.160	-0.408	-1.018	—
x3	L1	—	-0.572	0.783	-0.429	0.612	—
	L2	—	0.520	-0.004	0.369	-0.271	—
	L3	—	0.126	-0.063	-0.123	-0.002	—
x4	L1	—	—	1.300	0.186	-0.661	—
	L2	—	—	-0.171	-0.181	0.368	—
	L3	—	—	-0.149	0.020	-0.066	—
x5	L1	—	—	—	—	-0.006	—
	L2	—	—	—	—	-1.059	—
	L3	—	—	—	—	0.442	—

Table 8. Comparison of Multiple Correlation Coefficients Between Experienced and First-time Users

Kansei Dimension	Experienced	First-time	Difference (New – Exp)
Friendliness	0.341	0.723	+0.382
Convenience	0.472	0.769	+0.297
Thoughtfulness	0.287	0.756	+0.469
Immersion	0.441	0.618	+0.177
Authenticity	0.539	0.620	+0.081
Safety	0.628	—	—

4.4 Summary

The results indicate that while current FWT interfaces perform well in terms of functional usability, their effectiveness varies significantly across user groups.

Interface design factors show stronger influence on first-time users, whose emotional perceptions are highly dependent on information clarity, presentation, and guidance. In contrast, experienced users are less sensitive to design variations and rely more on efficiency and structured navigation.

These findings suggest that the primary challenge of FWT interface design lies not in improving functionality, but in enhancing emotional accessibility, particularly for first-time users.

ACKNOWLEDGMENTS

The author would like to express sincere gratitude to the academic mentors and colleagues who provided valuable guidance, constructive feedback, and continuous support throughout this research. Their insights greatly contributed to the development of this study and the preparation of this manuscript.

The author also appreciates all participants who contributed their time and expertise to this research.

REFERENCES

FREETOUR.com. (n.d.). *Free walking tours worldwide*. Retrieved May 15, 2026, from https://www.freetour.com

GuruWalk. (n.d.). *GuruWalk free walking tours*. Retrieved May 15, 2026, from https://www.guruwalk.com

Julia Kathryn Giddy & Gijsbert Hoogendoorn (2018) Ethical concerns around inner city walking tours, Urban Geography, 39:9, 1293-1299, DOI: 10.1080/02723638.2018.1446884

Like It Formosa. (n.d.). *Taiwan free walking tours*. Retrieved May 15, 2026, from https://www.likeitformosa.com

Markwell, K., Stevenson, D., & Rowe, D. (2004). Footprints in the city: Walking tours and the production of tourist space. In *Proceedings of tourism and hospitality research conference*.

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics, 15*(1), 3–11.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*.

SANDEMANs Tours. (n.d.). *SANDEMANs New Europe walking tours*. Retrieved May 15, 2026, from https://www.neweuropetours.eu

Yael Ram, Colin Michael Hall, (2018) "Walking tourism in cities: introducing the special issue", International Journal of Tourism Cities, Vol. 4 Issue: 3, pp.281-284, https://doi.org/10.1108/IJTC-09-2018-098

An Analysis of the Psychological Impacts of Collaborative Exploration Activities in Group Travel Contexts on Well-being

Kai YAMAGUCHI[1a], Akihiro OGINO [a]

[a] *Faculty of Information Science and Engineering, Kyoto Sangyo University, Kyoto, Japan, i2686325@cc.kyoto-su.ac.jp, ogino@cc.kyoto-su.ac.jp*

ABSTRACT

Group sightseeing sometimes fails to promote active interaction among participants and improve their well-being. This study investigated how collaborative exploration activities affect group tourism participants' well-being, using a photo-spot quiz as an experimental task to induce cooperative behavior. We developed a group tourist well-being support system using iPhones and Apple Watches that presents a partial photo of a tourist site as a task, verifies completion through photo similarity and GPS, and provides different hints to each participant to encourage discussion. A field experiment was conducted in Kyoto, Japan, with six groups of three visiting sightseeing locations across two conditions (with vs. without the system) on separate days. Personality traits were collected before the experiment; emotional states were collected in three phases (arrival, mid-visit, exit); movement distance and heart rate were recorded throughout the tour. Results showed that groups with more participants high in extraversion and openness exhibited increased positive affect during collaborative exploration, and their heart rate synchronization remained more stable even in highly crowded locations. These findings suggest that collaborative exploration can enhance group tourism experiences, depending on participant characteristics and environmental context.

Keywords: *Cooperative behavior, Photo-spot quiz, Personality traits, Heart Rate Synchronization*

1 INTRODUCTION

More than 80% of overnight trips in Japan are conducted in groups rather than individually (Recruit LifeStyle, 2025). However, simply being together at the same place does not necessarily guarantee active interaction or improved well-being among participants, and few studies have directly examined this relationship. Existing destination recommendation systems for group tourism (Alatiyyah, 2025) optimize which spots a group should visit but do not address how members can enjoy the experience together on-site.

Game-based tourism can influence participants' emotions and enhance engagement (Xu et al., 2017), but has not focused on improving on-site stays. Behavioral and physiological data

1 Corresponding author. E-mail: i2686325@cc.kyoto-su.ac.jp

such as movement and heart rate have been used to examine how engaging in activities together at a tourist spot affects enjoyment (Ogino et al., 2022), but without considering the personality traits of group members. To investigate how personality differences affect enjoyment and emotions during group exploration, we developed a well-being support system for group tourists using an iPhone and Apple Watch.

2 METHOD

2.1 System

We developed a group tourism well-being support system using iPhones and Apple Watches for a group of three participants. Each participant carried an iPhone, which recorded GPS-based routes, and wore an Apple Watch, which measured heart rate. The photo-spot quiz application was installed on only one of the iPhones because the system aimed to help the group work together through active conversations at tourist destinations (Figure 1). The application presented a partial photo of a tourist spot as an exploration task and encouraged participants to find the location together. Task completion was judged based on similarity between the participants' photo and the correct photo, computed with a machine-learning library (Core ML) on the iPhone, together with GPS information to verify the location. The app provided three types of hints; to facilitate group discussion, each participant could only see hints different from the others, so they needed to share information and collaborate to solve the task. The application also provided a detailed screen for each tourist spot (Figure 2). From this screen, participants can check the quiz photo, access questionnaires, view hints, and verify the photos they took.

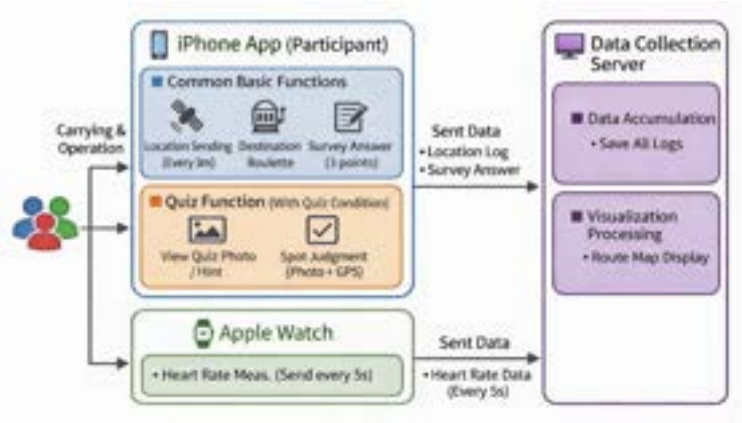


Figure 1. Overview of the group tourism well-being support system



Figure 2. Example of the photo-spot quiz application interface

2.2 Participants

A total of 17 participants (10 males and 7 females; mean age = 21.67 years, SD = 1.38) participated in the experiment. They were divided into six groups, each consisting of three close friends. However, one participant was included in two different groups.

2.3 Experiment Procedure

We conducted the experiment over two days to evaluate whether the photo-spot quiz application made group sightseeing more fun and enjoyable for the participants. Each group of three conducted two conditions of the experiment: sightseeing with and without the photo-spot quiz application. In the quiz application condition, the participants collaboratively performed a photo-spot quiz task while sightseeing. In contrast, in the non-quiz application condition, the participants explored tourist spots without performing any collaborative tasks.Each condition was tested on separate days (Figure 3). The group visited three tourist spots in one day. They chose these spots from a list of 12 in Kyoto, Japan. The participants in a group discussed and decided together where to go before each day's tour. Importantly, the three spots visited on the day with the system and the three spots visited on the day without the system were entirely different even for the same group, because each group newly selected the spots from the list of 12 before each day's tour. This design was adopted to avoid carryover effects from revisiting identical spots. At each location, they filled out surveys three times: upon arrival, at a mid-visit checkpoint, and upon departure. The mid-visit checkpoint was defined differently across conditions—immediately after quiz completion in the with-system condition, and seven minutes after arrival in the without-system condition—a design choice discussed further in the Limitations section. During the experiment, each participant used their iPhone to track their route and moving distance during their tour using the GPS. They also recorded their heart rates using Apple Watches while sightseeing. These measurements were taken in both situations to compare the emotional, behavioral, and physical responses.

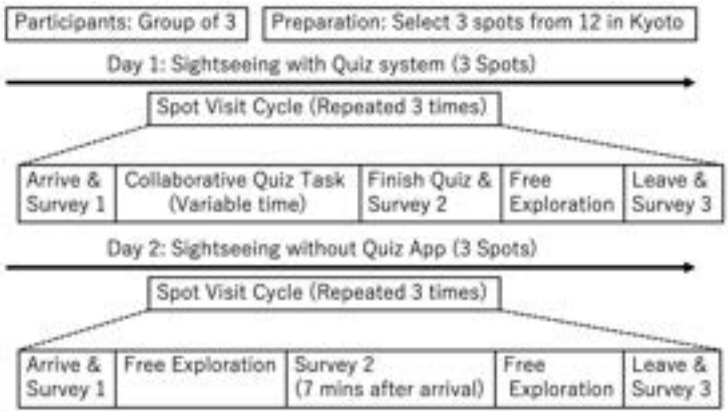


Figure 3. Overview of the Experimental Procedure

2.4 Measures

2.4.1 Emotional Measures

Emotional changes were assessed with the Multidimensional Mood Scale (MMS; Terasaki et al., 1992) and the Japanese version of the Positive and Negative Affect Schedule (PANAS; Sato et al., 2001). The MMS comprises eight mood scales (Depression/Anxiety, Hostility, Boredom, Liveliness, Well Being, Friendliness, Concentration, Startle), each with five items rated on a four-point Likert scale (score range: 5–20). The PANAS measures positive affect (PA) and negative affect (NA), each with eight items on a six-point Likert scale (score range: 8–48). At the exit survey, participants additionally rated perceived crowding (1–5) and quiz difficulty (1–4).

2.4.2 Personality Traits

Participants' personality traits were assessed prior to the experiment using a questionnaire based on the Big Five Personality Model. The five dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness) were measured to examine how individual differences influenced the effectiveness of collaborative exploration during group tourism.

2.4.3 Behavioral Measures

Behavioral data were collected using GPS data obtained from the participants' iPhone devices. During sightseeing, GPS data were transmitted from the iPhone application to a server each time a participant moved more than three meters. This approach enabled the continuous tracking of participants' locations while minimizing redundant data transmission. Using the collected GPS data, movement routes and total movement distances within each tourist spot were recorded during the experiment.

2.4.4 Physiological Measures

Physiological data were collected by continuously measuring the participants' heart rates using Apple Watch devices during sightseeing. Heart rate data were recorded throughout the experiment under both quiz and non-quiz conditions. The collected heart rate data were used to examine changes in physiological arousal during sightseeing activities. In addition, heart rate synchronization among group members was analyzed as an indicator of shared engagement and

coordinated activity during the collaborative exploration, following previous research on group mood estimation using wearable devices (Ogino & Ikematsu, 2022). Heart rate synchronization was evaluated by examining the temporal similarities in heart rate fluctuations among group members at each tourist spot.

3 RESULTS

3.1 Group Classification Based on Personality Traits

Non-metric multidimensional scaling (NMDS) was applied to the Big Five Questionnaire data (Wada, 1996) to examine similarities and differences in participants' personality traits. Figure 4 shows the participants in NMDS space, and Table 1 shows how each NMDS axis relates to the Big Five traits. Participants in the upper region scored higher on extraversion and openness; those in the lower region scored lower on these traits.

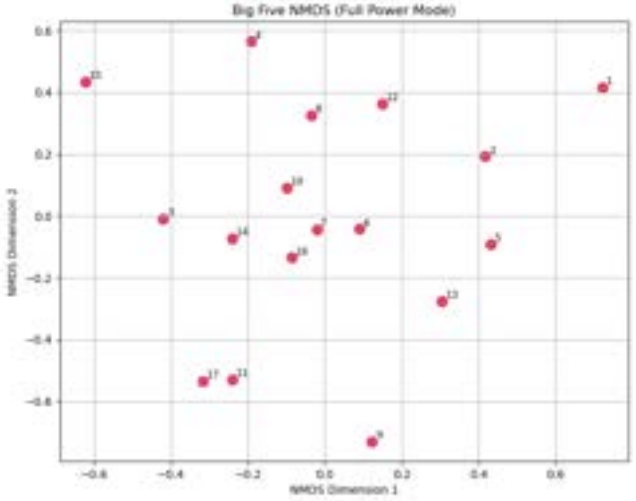


Figure 4. NMDS plot of participants based on Big Five personality traits.

Table 1. Correlations between NMDS dimensions and Big Five personality traits

	Extraversion	Neuroticism	Openness	Conscientiousness	Agreeableness
NMDS1 (Horizontal axis)	-0.35	0.82	-0.50	-0.51	0.21
NMDS2 (Vertical axis)	0.86	0.33	0.58	-0.27	0.42

Participants were divided into two groups. Group A consisted of groups with two or more of three members relatively high in extraversion and openness, and Group B was defined by the opposite pattern. In small groups of three, the personality traits shared by most members are

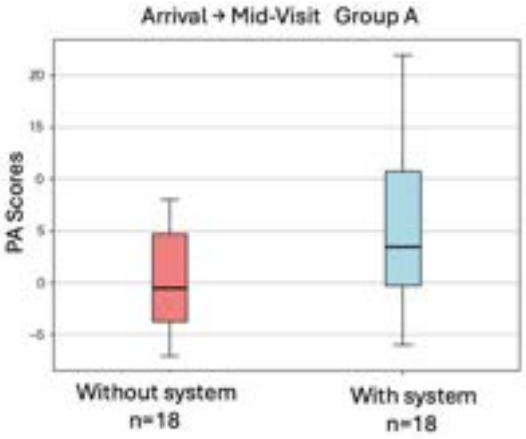
likely to shape how the group behaves, interacts, and feels during sightseeing. All subsequent analyses compared Group A and Group B.

3.2 Effects of Collaborative Exploration on Emotional States Across Personality Groups

Change scores in PANAS and MMS were calculated across three intervals (Arrival→Mid-visit, Mid-visit→Exit, Arrival→Exit) and compared between conditions using the Wilcoxon signed-rank test ($p < 0.05$), visualized with box plots.

As shown in Figure 5, in Group A, PA scores significantly increased from the arrival to the mid-visit phase under the photo-spot quiz application condition (With system) compared with the non-quiz application condition (Without system). Similarly, the “Liveliness” scores were significantly higher in the quiz application condition than in the non-quiz application condition. This indicates that the participants experienced a more energetic and active emotional state during collaborative exploration. In contrast, no significant changes were observed in Group B.

In contrast, no significant changes were observed in Group B. Furthermore, in Group A, NA scores significantly increased from arrival to exit under the "With system" condition compared with the "Without system" condition. A closer inspection of the individual NA scores revealed that this increase was primarily driven by higher scores on the items “worried” and “upset.” Regarding the "startle" item, a significant increase from arrival to mid-visit was observed in the photo-spot quiz condition. This increase was consistently observed across groups, regardless of personality composition. In contrast, no comparable increase was observed in the "Without system" condition.



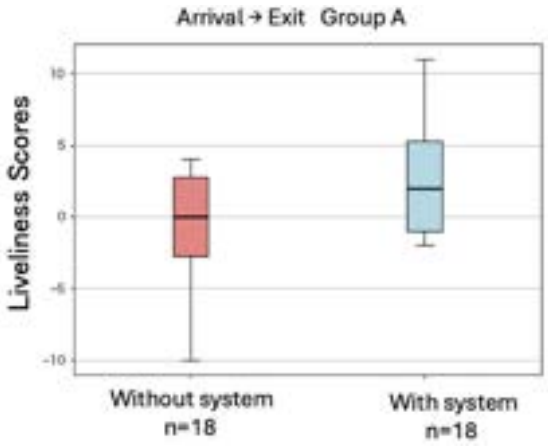


Figure 5. Changes in PA scores (left) in PANAS and Liveliness scores (right) in the MMS from Arrival to Mid-Visit in Group A

Furthermore, in Group A, NA scores significantly increased from arrival to exit under the "With system" condition compared with the "Without system" condition. A closer inspection of the individual NA scores revealed that this increase was primarily driven by higher scores on the items “worried” and “upset.” Regarding the "startle" item, a significant increase from arrival to mid-visit was observed in the photo-spot quiz condition. This increase was consistently observed across groups, regardless of personality composition. In contrast, no comparable increase was observed in the "Without system" condition.

3.3 Effects of Perceived Crowding on Emotional Changes

Tourist spots were categorized into high- and low-crowding locations based on questionnaire responses, and emotional changes were compared using the Mann-Whitney U test.

As shown in Figure 6, during the Arrival to Mid-Visit interval at high-crowding spots, the “Liveliness” scores were significantly higher under the "With system" than under the "Without system" ($p = 0.049 < 0.05$). This pattern indicates that collaborative exploration enhances energetic and active emotional states, even in crowded environments.

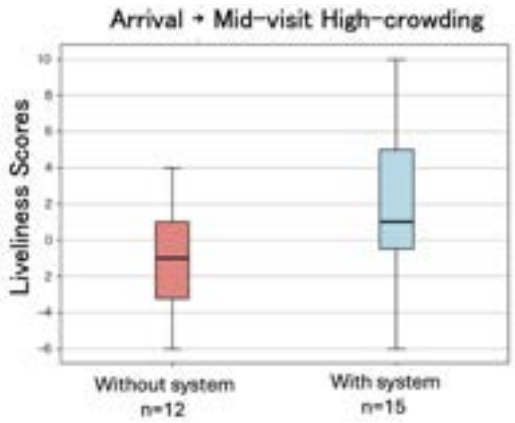


Figure 6. Change in Liveliness Scores from Arrival to Mid-Visit at High-Crowding Spots

Table 2. Mean and Standard Deviation of Heart Rate Synchronization Scores for Each Group

	Condition	SD	Mean
G1	Without System	0.11	0.52
	With System	0.07	0.40
G2	Without System	0.24	0.45
	With System	0.06	0.60
G3	Without System	0.10	0.58
	With System	0.08	0.48
G4	Without System	0.06	0.34

	With System	0.09	0.39
G5	Without System	0.31	0.29
	With System	0.03	0.62
G6	Without System	0.12	0.62
	With System	0.02	0.54

3.4 Heart Rate Synchronization

Heart rate synchronization was quantified as the number of matched heart-rate peaks among group members within a ±5 s window, using standardized signals from Apple Watches (sampled every 5 s).

Table 2 presents the mean values and standard deviations of the heart rate synchronization scores for each group under both the "With system" and "Without system" conditions. No remarkable differences in mean values were observed between the conditions. However, the standard deviations tended to be smaller under the "With system" condition, suggesting that synchronization was more stable during collaborative exploration.

4 DISCUSSION

Groups with more participants high in extraversion and openness showed significantly greater increases in Positive Affect (arrival to mid-visit) and Liveliness (arrival to exit), while groups lower in these traits did not. This shows that the photo-spot quiz application works well-being for people who are more "extraversion" and "openness." These people like social activities and new experiences, which might be why they enjoy the quiz more. This interpretation is consistent with prior work showing that extraversion is dispositionally associated with heightened positive affect (Costa & McCrae, 1980), and that openness predicts a preference for novel and exploratory experiences (McCrae, 1996). An alternative account, however, is that individuals high in extraversion report higher positive affect in general (Costa & McCrae, 1980). This baseline tendency could contribute to the Group A effect independently of the task; distinguishing these accounts would require designs that separate trait-level baseline affect from task-induced changes. However, they also increased the "Negative affect" score by increasing the scores in the “worried” and “upset” items. This indicates that temporary tension was experienced while searching for the correct location during the task.

The result indicated that the environmental context also affects people using the photo-spot quiz application. In very crowded places, people felt more "Liveliness" from the Arrival phase to the Mid-Visit phase when using the photo-spot quiz application. In less crowded places, there was no significant change. This implies that collaboration on tasks can be highly effective in crowded environments. The tasks provide people with a clear goal, making the crowded environment feel like part of the activity rather than a source of stress. This interpretation is consistent with the attentional focusing that characterizes flow states, in which engagement with a goal-directed task reduces the perceived cost of environmental stressors (Csikszentmihalyi, 1990). Importantly, the same crowded spots did not produce a comparable

rise in emotional state under the without-system condition, which argues against a purely site-attractiveness explanation and supports the interaction between crowding and the collaborative task. An alternative account is that crowded spots may make the quiz itself more difficult, so that a stronger sense of achievement upon task completion contributes to the observed rise in Liveliness. In less crowded places, people can move around easily and enjoy the view, which lowers their stress. In these places, the photo-spot quiz may not be as effective. To improve the experience in these places, it may be more helpful to focus on ways to enjoy the scenery more deeply or improve the quality of the experience rather than on activities with strict goals.

"Startle" scores increased significantly during the Arrival to Mid-Visit phase only under the with-system condition, likely reflecting the moment participants found the answer and enjoyed it.

5 LIMITATIONS

5.1 Sample size and non-independent observations

The study included 17 participants across six groups of three, and one participant took part in two different groups. Both of these groups were classified as Group B, so no participant contributed data to both sides of the Group A versus Group B comparison, and that comparison is not confounded by the repeated participant. Their responses do, however, carry slightly more weight within the Group B statistics. The within-condition Wilcoxon signed-rank comparisons are unaffected, since each participant serves as their own control across conditions. Given the small number of groups, we treat these findings as preliminary and await replication with a larger, fully independent sample.

5.2 Choice of personality inventory

We measured the Big Five using Wada's (1996) scale, a standard Japanese-language instrument that keeps our results comparable with earlier Japanese personality research. Item loadings can vary frozm one Big Five scale to another, so administering a second inventory in parallel would give a firmer basis for the group classification we relied on.

5.3 Unmeasured within-group relational characteristics

Each group consisted of three close friends, but we did not record how close they were, or in what way. Length of acquaintance, frequency of interaction, prior shared travel, and whether one member tended to lead were all left uncollected. Two groups that both count as “close friends” could differ on any of these, and that hidden variation may have added noise to the outcomes. Including a validated closeness scale in future rounds would let us treat these factors as moderators rather than leaving them uncontrolled.

5.4 Design-driven timing difference in the mid-visit questionnaire

The mid-visit questionnaire was administered at different moments in the two conditions: immediately after quiz completion in the with-system condition, and seven minutes after arrival in the without-system condition. This was intentional. In the with-system condition we wanted to capture the emotional state right after task completion, and fixing a specific elapsed time

would have interrupted many groups mid-task. The without-system condition has no equivalent task-completion event, so we used a fixed elapsed time to approximate the moment after an initial round of free exploration. The gap between the two was nevertheless large—10.3 ± 2.4 minutes without-system versus 21.8 ± 13.2 minutes with-system (Wilcoxon signed-rank test, $p = 0.003$)—so part of the increase in Positive Affect and Liveliness may reflect longer time spent at the location rather than collaborative exploration itself. Adding a few time-based checkpoints alongside the functional one used here would help separate the two.

5.5 Physical versus psychological contributions to heart rate synchronization

Heart rate synchronization was used as an indicator of shared engagement, but heart rate reflects physical activity as much as psychological state. The standardization procedure applied here does not separate synchronization arising from shared movement (e.g., walking together) from synchronization arising from shared psychological engagement with the task. The results should be read as a combination of both. A physiological signal less confounded by movement—electrodermal activity, for example—would help isolate the psychological component in future work.

6 CONCLUSION

This study investigated how a photo-spot quiz system affects well-being in group tourism experiences. The system was most effective for groups higher in extraversion and openness, who reported greater positive affect and liveliness during collaborative exploration. In crowded environments, the system enhanced emotional engagement. Physiologically, heart rate synchronization remained more stable under the system across most groups. Together, these findings suggest that collaborative exploration can enhance group tourism experiences, contingent on participant personality and environmental context.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 23K11294.

REFERENCES

Recruit LifeStyle Co., Ltd. (2025). *Jalan Travel Survey: Domestic Lodging Travel Survey* (in Japanese).

Alatiyyah, M. (2025). A novel group tour trip recommender model for personalized travel systems. PeerJ Computer Science, 11, e2589.

Xu, F., Buhalis, D., & Weber, J. (2017). Serious games and the gamification of tourism. *Tourism Management*, 60, 244–256.

Wada, S. (1996). Construction of the Big Five scales using personality trait terms. *The Japanese Journal of Psychology*, 67(1), 61–67.(in Japanese)

Terasaki, M., Kishimoto, Y., & Koga, A. (1992). Construction of a multidimensional mood scale. *The Japanese Journal of Psychology*, 62(6), 350–356. (in Japanese)

Sato, T., & Yasuda, A. (2001). Development of the Japanese version of the Positive and Negative Affect Schedule (PANAS). *The Japanese Journal of Psychology*, 72(5), 400–408. (in Japanese)

Ogino, A., & Ikematsu, Y. (2022). Mood Estimation Method in a Group Using Smartwatches to Support Positive Tourism Experiences. International Journal of Affective Engineering, 21(1), 33–42.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. Journal of Personality and Social Psychology, 38(4), 668–678.

McCrae, R. R. (1996). Social consequences of experiential openness. Psychological Bulletin, 120(3), 323–337.

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.

BRIDGING EMOTIONAL PREFERENCE AND ENVIRONMENTAL ASPIRATION IN DODOL PACKAGING DESIGN THROUGH KANSEI ENGINEERING

Novi Purnama Sari[1a], Wiwi Prastiwinarti[a], Anggi Nuraini[b] , Aisyah Nur Zahirah[c]

[1a] *Jakarta State Polytechnic, Indonesia, novi.purnamasari@grafika.pnj.ac.id*
[a] *Jakarta State Polytechnic, Indonesia, wiwi.prastiwinarti@grafika.pnj.ac.id*
[b] *Jakarta State Polytechnic, Indonesia, anggi.nuraini.tgp23@stu.pnj.ac.id*
[c] *Jakarta State Polytechnic, Indonesia, aisyah.nur.zahirah.tgp23@stu.pnj.ac.id*

ABSTRACT

Dodol, a traditional Indonesian food product, faces persistent packaging challenges related to perceived safety, visual appeal, and environmental considerations. This study aims to bridge the gap between consumers' emotional preferences and environmental aspirations in Dodol packaging design through a Kansei Engineering-based approach. Consumer emotional responses were collected using Kansei words and analyzed through Principal Component Analysis (PCA) to identify dominant design concepts. Specific design elements were determined using the Quantification Theory Type I (QTT1) method. The results indicate that aluminum foil emerged as the preferred primary packaging material due to its strong emotional association with hygiene, safety, and product quality, while environmentally oriented alternatives such as commercial edible packaging were perceived as unfamiliar, expensive, and difficult to implement in small-scale industries. To address this gap, secondary packaging was strategically developed based on the Classic–Trendy design concept with art carton material, a rounded rectangular shape, a secure closure lock and drawer feature, a photo illustration, a full color composition, and an informative label. The environmental performance in this study is discussed from a design-oriented perspective and consumer perception, rather than from a quantitative environmental impact assessment. These findings demonstrate how secondary packaging design can mediate emotional acceptance and support environmental aspirations in traditional food packaging through Kansei Engineering.

Keywords: *Kansei Engineering; Packaging Design; Environmental Performance; Emotional Preference; Traditional Food Packaging*

1 INTRODUCTION

Local culinary heritage contributes significantly to cultural identity and sustainable economic development (Chukwurah et al., 2025). Dodol is a sweet treat recognized as a cultural heritage in several Southeast Asian countries, such as Indonesia, Singapore, Malaysia, Thailand, and the Philippines (Yeoh et al., 2024). In Indonesia, Dodol is an alternative souvenir product, but it's packaging often lacks attention to safety, visual appeal, and environmental considerations.

[1] Corresponding author.

One example is Dodol Picnic, a Garut specialty, which has been around since 1956 and has earned the Wonderful Indonesia Co-Branding award. Despite its strong brand and product quality, the design of its secondary and primary packaging still fails to optimally meet consumers' emotional needs. According to Yu et al (2024), strengthening brand identity, using more engaging visuals, and presenting information are crucial to increasing consumer interest and trust. These design elements serve as communication tools that convey the product's character. Furthermore, paper-based primary packaging often sticks to the product, causing a risk of ingestion, as shown in Figure 1. This can trigger the migration of chemicals in the ink, reducing food safety, and increasing the risk of health impacts (Seref & Cufaoglu, 2025). Beyond its protective function, packaging plays a crucial role as a communication tool. Research by Ooijen et al. (2017) found that packaging design elements influence conclusions about product quality. Meanwhile, research by Khan & Lee (2020) highlights packaging function as a communication medium that uses a universal language, proven to shape consumer perceptions across demographic groups.



Figure 1. Dodol Picnic product packaging and primary paper packaging that is often eaten

According to Sostar et al. (2025), research on consumer behavior in recent years has increasingly emphasized psychological and emotional factors that influence purchasing decisions. Two key components that significantly influence consumer perceptions of quality and preferences are packaging and branding. According to Lewis et al. (2014), emotions are the primary drivers influencing environmentally friendly purchasing decisions. This study examines consumers' emotional and rational evaluations of environmentally friendly packaging. The conceptual model combines individuals' general environmental concerns, their rational beliefs about the environmental impact of product consumption, and the emotions they evoke.

According to Sostar et al. (2025), the food and beverage industry faces the challenge of balancing consumer preferences for packaging with environmental responsibility. Packaging materials are generally associated with packaging sustainability and waste reduction. According to Atta-Delgado et al. (2024), despite increasing environmental awareness in packaging design, consumer acceptance of environmentally oriented packaging remains heavily influenced by emotional factors. However, these emotional drivers do not always align with environmentally friendly materials, creating considerations that require careful examination. This creates a gap between emotional preferences and environmental aspirations, particularly for traditional food packaging. Research focusing on comparing product acceptance from a design-oriented perspective and consumer perception is particularly interesting. According to Bou-Mitri et al. (2021), emotional responses related to perceived cleanliness, safety, and quality significantly influence consumer preferences. Kansei engineering provides an effective framework for exploring these considerations by revealing how consumers emotionally evaluate alternative packaging solutions (Gao et al., 2024).

Kansei Engineering integrates emotional responses into packaging development (Schütte et al., 2024). The advantage of Kansei Engineering lies in its ability to identify consumer needs through emotional and psychological aspects (Soenandi et al., 2021). A combination of Principal Component Analysis (PCA) and Quantification Theory Type 1 (QTT-1) methods is used to support the packaging development process. According to Ahmad et al. (2025), PCA helps interpret the emotional relationship of Kansei words more clearly based on respondents' responses. The QTT-1 method is used to analyze the relationship between packaging elements and the desired packaging concept (Nagamachi & Lokman, 2015). This study aims to bridge the gap between consumers' emotional preferences and environmental aspirations in Dodol packaging design through a Kansei Engineering-based approach.

2 METHOD

The research flow is shown in Figure 2 below.

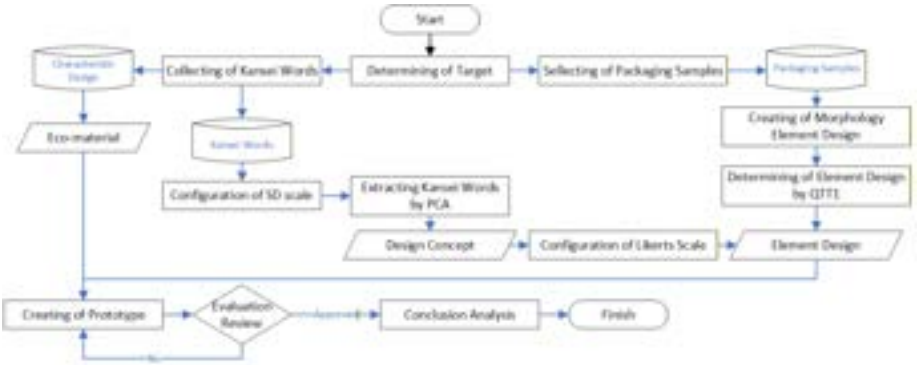


Figure 2. Process flow chart

The initial stage of collecting Kansei words from emotional respondents used an open-ended questionnaire with packaging sample stimuli (Mouludi & Sipahutar, 2024). Respondents were selected using purposive judgment sampling based on the criteria that they had previously consumed Dodol products and were familiar with packaged traditional foods. Participants consisted of males and females aged between 18 and 40 years totaling at least 30 respondents. According to Eltohamy et al. (2022), the sample size (n) was 30 based on the Central Limit Theorem (CLT), which shows a near-normal distribution of the mean for the appropriate parametric test. The obtained Kansei words were then identified as adjectives as Kansei words and non-adjectives as design characteristics (Chang et al., 2017). According to Nagamachi (2011), packaging samples were collected via the internet, with a minimum of 20-25 packaging samples and a minimum number of Kansei words of 30. The samples used were samples relevant to Dodol products (Papantonopoulos et al., 2021). The selection process was carried out by looking at similarities in terms of material, shape, design style, and size (Sari et al., 2024). According to Lin et al. (2022) , the obtained Kansei words were then examined for correlation with the selected samples using a semantic differential (SD) questionnaire with a 7-point scale. This scale is considered to have a good level of sensitivity. The questionnaire data results were then processed to determine Kansei word clusters as design concepts using the PCA method. According to Hasan & Abdulazeez (2021), this method is believed to be effective in reducing the dimensionality of

large data without losing information. The design concept will be assessed for correlation with the samples using a Likert questionnaire (Zuo & Wang, 2020). On the other hand, packaging samples will be identified based on the morphology of their design elements (Čok et al., 2022). The results of morphological identification and respondent perception data are synthesized using the QTT1 method (Zuo & Wang, 2020). This process will produce the most significant packaging design elements based on the design concept. The final stage is the creation of a prototype based on the results obtained and compared with alternative packaging prototypes that consider environmentally friendly materials. The results of this evaluation will indicate consumer acceptance preferences for packaging designs that prioritize emotional preferences or environmental aspirations.

3 RESULT AND DISCUSSION

3.1 Design Concept of Packaging

Initially, 80 packaging samples were obtained, which were then narrowed down to 75 distinct samples, as shown in Figure 3.



Figure 3. Packaging samples

Table 1 shows examples of Kansei words (adjectives) and design characteristics (non-adjectives) with the number of times they were used by respondents. A total of 44 Kansei word pairs and 24 design characteristics were selected from a pool of 328 words based on the emotional responses of the first 40 respondents.

Table 1. The example of Kansei words and design characteristics

Kansei Word	QTY	Design Characteristic	QTY
Design interprets “sweet”	34	Brown-colored design	32
Non-sticky primary packaging	33	Aluminum foil primary packaging	11
Design interprets “chewy”	25	Plastic primary packaging	10
.....	....	.....	....
.....	....	.....	....
Innovative	1	Handle feature	1
Modern	1	Window feature	1

The results of the PCA analysis are shown in Figure 4. According to Jollife & Cadima (2016), a high standard deviation value indicates that the variable or Kansei word has good quality to be retained. According to Ma et al. (2021), the principal components are retained if they have a

Cumulative Proportion of Variance of at least 80%. PC1 and PC2 have Cumulative Proportion of Variance standard deviation values of (8.3928; 93.92%) and (1.12336; 95.6%).

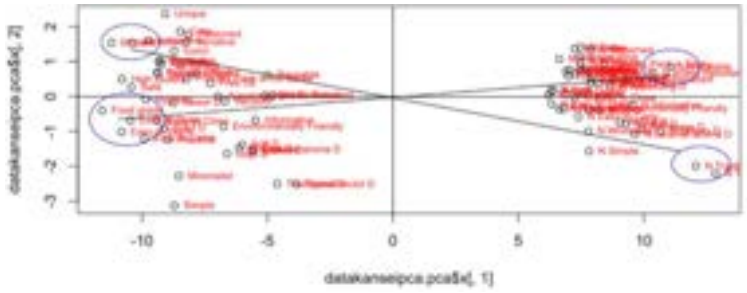


Figure 4. Plotting of Kansei word distribution using PCA

The concept is arranged based on a collection of Kansei words from each Main Component (PC) with the highest positive and negative values, or those at the very end of Figure 4. The positive side of PC 1 is the word "packaging that cannot be reused after being opened". In contrast, the negative side is the word "food-grade packaging that is easy to open and close, practical for storage, and easy to carry" so that the concept is concluded with the expert panelists' "cannot be reused" - "functional". The positive side of PC2 is the word "modern, attractive, and innovative design", while the negative side is the word "generic packaging appearance that reduces the unique characteristics of Dodol", so the concept is "trendy" - "classic".

3.2 Element Design Packaging

The morphological elements of the packaging sample in Figure 3 were identified through discussions with expert panelists in the field of packaging materials and design, as shown in Figure 5.

Element Design	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Element Design	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Type 10	Type 11	Type 12
X1 Material primer	(GR, X1,1) Greaseproof paper	(GL, X1,2) Glassine paper	(I, X1,3) Ivory	(KR, X1,4) Kraft	(PP, X1,5) Polyprop ylene plastic	(MLP, X1,6) Multi layer plastic	(BP, X1,7) Blister pack	(AL, X1,8) Aluminium foil	(N, X1,9) Nothing	X2 Material sekunder	(I, X2,1) Ivory	(AC, X2,2) Art cartoon	(D, X2,3) Duplex	(CB, X2,4) Corrugated box	(PB, X2,5) Paperboard	(KR, X2,6) Kraft	(MLP, X2,7) Multiblayer paper	(PP, X2,8) Polyprop ylene plastic	(S, X2,9) Steel	(R, X2,10) Rattan		
X3 Design style	(N, X3,1) Natural	(F, X3,2) Fun	(T, X3,3) Traditional	(M, X3,4) Modern	(P, X3,5) Premium	(E, X3,6) Elegant	(Mi, X3,7) Minimalist	(S, X3,8) Simple		X4 Form primer	(S, X4,1) Sphere	(SC, X4,2) Square Cube	(R, X4,3) Rectangle	(H, X4,4) Hexagon	(T, X4,5) Triangle	(PY, X4,6) Pyramid	(PL, X4,7) Polygonal	(P, X4,8) Pouch	(C, X4,9) Candy	(U, X4,10) Umbrella	(N, X4,11) Nothing	
Element Design	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Type 10	Type 11	Type 12	Type 13	Type 14	Type 15	Type 16	Type 17	Type 18	Type 19	Type 20		
X5 Form sekunder	(R, X5,1) Rectangle	(FR, X5,2) Flattened rectangle	(VR, X5,3) Vertical rectangle	(SC, X5,4) Square Cube	(FS, X5,5) Flattened Square	(RS, X5,6) Rounded square	(RFS, X5,7) Rounded flattened square	(P, X5,8) Pentagon	(H, X5,9) Hexagon	(T, X5,10) Polygonal Tube	(GS, X5,11) Gem shape	(ATT, X5,12) Arch top tube	(C, X5,13) Cylinder	(PR, X5,14) Prism	(PU, X5,15) Purse	(B, X5,16) Bag	(SP, X5,17) Standing pouch	(FP, X5,18) Floral shape	(OS, X5,19) Organic shape	(CM, X5,20) Custom model		
Element Design	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Element Design	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8	Type 9	Type 10	Type 11	
X6 Lid	(FT, X6,1) Flip Top	(TF, X6,2) Top Flap	(FTF, X6,3) Front Tuck Flap	(SF, X6,4) Special Flap	(EF, X6,5) Envelope Fold	(SL, X6,6) Slide lid	(FL, X6,7) Friction lock	(TP, X6,8) Two piece	(S, X6,9) Sealer	X7 Feature	(T, X7,1) Tray	(TW, X7,2) Tray & Window	(RW, X7,3) Rectangle window	(CW, X7,4) Circle window	(CUW, X7,5) Custom window	(DRW, X7,6) Drawer & Rectangle window	(DCW, X7,7) Drawer & Circle window	(D, X7,8) Drawer	(DD, X7,9) Dispenser drawer	(H, X7,10) Handle	(R, X7,11) Ribbon	(NF, X7,12) No feature
Element Design	Type 1	Type 2	Type 3	Element Design	Type 1	Type 2	Type 3	Type 4	Element Design	Type 1	Type 2	Type 3	Element Design	Type 1	Type 2	Type 3						
X8 Illustration	(P, X8,1) Photo	(I, X8,2) Illustration	(N, X8,3) Nothing	X9 Color	(FC, X9,1) Full colour	(W, X9,2) Warm	(N, X9,3) Neutral	(C, X9,4) Cool	X10 Printing technique	(NP, X10,1) No print	(DP, X10,2) Direct printing	(IP, X10,3) Indirect printing	X11 Information	(I, X11,1) Informative	(SI, X11,2) Semi informative	(U, X11,2) Uninformative						

Figure 5. The morphological elements design

The morphological results showed that there were 11 factors from X1-X11 with types from 1-20. Each sample in Figure 3 is analyzed by factor category and type according to the

morphological elements in Figure 5. For example, Sample B has a primary material of multi-layer plastic (6), secondary material is ivory (1), design style is natural (2), primary shape is square cube (2), secondary shape is prism (14), cover is slide (6), feature is drawer (8), has illustration (2), full color (1), printing technique is direct (2), and semi-informational (2). Meanwhile, the emotional results of respondents' perceptions from the Likert questionnaire had an average of 5.5 are shown in Table 2 below. The data is used to determine packaging design elements using the QTT1 method. Based on the R-square value from running QTT1, the Classic-Trendy concept (0.9994) has a greater value than Unreusable-Functional (0.9990). Figure 6 shows the design element results for the Classic-Trendy concept.

Table 2. Analysis of sample packaging elements and the results of respondents' emotional perceptions of the concept

Sample	Design Element											Classic-Trendy Value	
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	Mean	Std. Dev
A	5	1	2	3	2	1	12	2	1	2	1	6.7	0.70
B	6	1	2	2	14	6	8	2	1	2	2	5.5	1.26
----------	----	----	----	----	----	----	----	----	----	----	----	----------	----------
VVV	5	5	3	2	5	8	12	2	4	1	3	6.1	1.47
WWW	1	2	2	3	2	2	12	2	3	2	2	6.7	0.94

According to Sari et al. (2020), the selection of design elements is based on the longest histogram bar of each factor because it has the greatest influence on design perception.

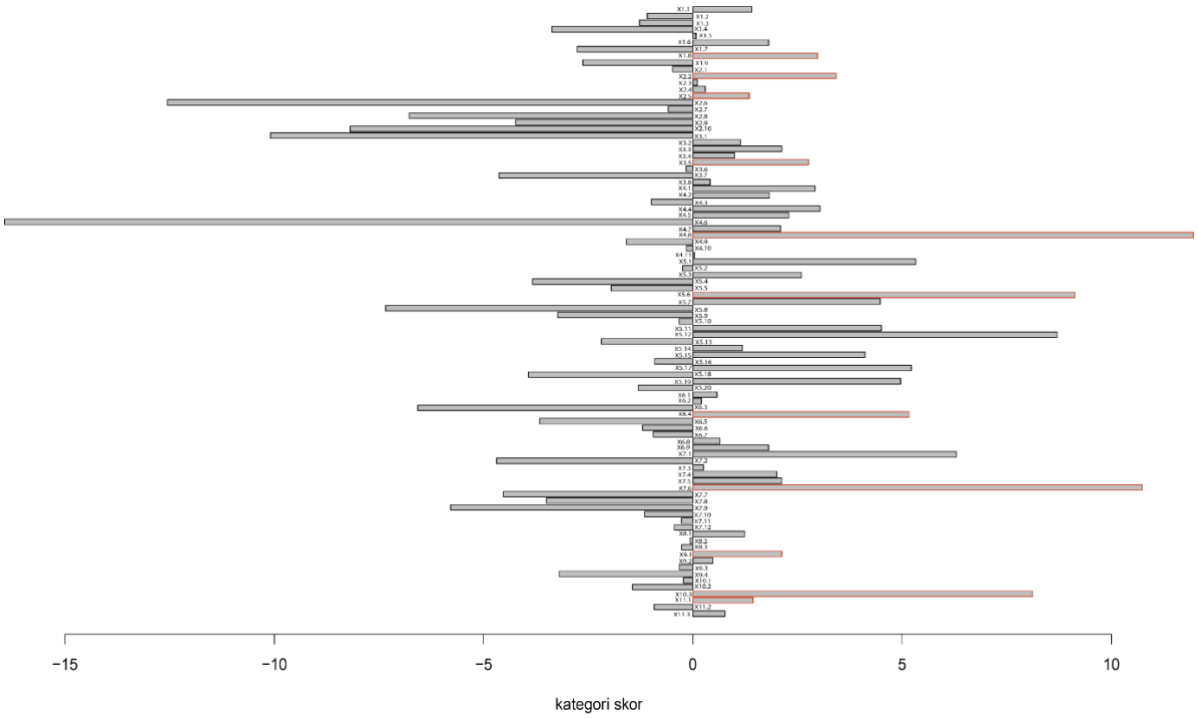


Figure 6. Design element results by QTT1

The 'Trendy' design element obtained primary packaging aluminum foil (X1.8), secondary material art carton (X2.2), premium design style (X3.5), primary packaging shape pouch (X4.8), secondary packaging shape round square (X5.6), special flap lock shape (X6.4), additional features

rectangular drawer (X7.6), photo illustration (X8.1), full color (X9.1), indirect printing technique (X10.3), and informative label design (X11.1).

3.3 Prototype

The resulting design elements align with the established concept, namely Trendy. The primary packaging material is aluminum foil (X1.8). This aligns with consumer emotions, as observed in the design characteristics in Table 1. Aluminum foil is the priority of material choice. According to Nugroho & Redjeki (2015), aluminum foil is considered aesthetic, safe, and hygienic. The secondary packaging material is art cardboard (X2.2), with the addition of a matte lamination touch, giving the product a trendy impression. The design style is achieved with a premium design style (X3.5) through the selection of serif fonts, namely "Bilderberg and TT Ramillas", which support a premium impression. According to Parra (2025), serif fonts are often used to support a premium impression. The gold color of the font also reinforces the premium impression (Sari et al., 2020). The main packaging form is a pouch (X4.8). This form is a new alternative to the main Dodol packaging. The pouch shape gives a trendy or modern impression (Wibowo et al., 2025). While the secondary packaging shape is a rounded square (X5.6), because the material chosen is art carton, the shape that can be implemented is close to a square. The chosen closing system is a special flap (X6.4). An additional feature is a rectangular drawer (X7.6). The use of photo illustrations (X8.1) in the packaging design can support a premium impression. Full color (X9.1) is implemented with a combination of pattern designs for a stronger premium style; the choice of dominant brown color on the outside is supported by the results of consumer emotions in the characteristic design in Table 1. This color is identical to the color of the Dodol product, which gives a distinctive impression. The combination of gold and minimalist design maintains the premium image of the packaging. The printing technique used is indirect printing (X10.3), and an informative design label (X11.1). All design elements are realized in the form of a 3D mockup as the final representation of the Trendy Dodol packaging concept using Adobe Illustrator and Blender. The mockup results for the trend concept are shown in Figure 7.



Figure 7. Prototype 3D

3.4 Evaluation of Environmental Performance

The final step is to evaluate the prototype results in Figure 7 against consumer perceptions. This research gap is addressed by measuring consumer perceptions of packaging design based on consumer emotional preferences and environmental aspirations through a Kansei Engineering-based approach. Packaging material is a key factor in supporting environmental sustainability. Eco-friendly packaging is highly valued. The results of this study indicate that the secondary packaging material is Art Carton, while the primary packaging material is aluminum foil. Research (Otto et al., 2021) indicates that natural assessments of consumer perceptions of environmental sustainability of paper/cardboard and metal materials are more in line with their impact on environmental sustainability. Eco-friendly packaging was simulated by substituting edible paper

for the primary Dodol packaging. This refers to the results of design characteristics where respondents proposed edible packaging. The survey was conducted with 20 respondents who had consumed Dodol using edible paper. The results showed a significant difference in consumer perceptions of aluminum foil and edible paper, with a significance level of 0.034 (<0.05) and the highest mean value for edible paper at 25.65. Therefore, it can be concluded that, based on consumer perception, edible paper has a higher environmental performance. However, consumer trust and a lack of understanding about edible packaging led to respondents' initial hesitation about the product. Consumers questioned the product's hygiene and the dangers of edible paper. This finding is consistent with research (Atta-Delgado et al., 2024), which showed that biodegradable packaging is widely accepted by respondents. However, the negative health impacts and lack of willingness to consume edible packaging reflect greater concerns about what consumers eat rather than the amount of food wasted. The solution is to provide clear information about the safety and benefits of packaging (Atta-Delgado et al., 2024). Consumers' perception of primary aluminum foil packaging conveys a premium, safe, and hygienic impression. This, in turn, creates consumer uncertainty in packaging choices. If consumers consider emotional environmental concerns, they will still choose recyclable packaging, but if they consider safety, they will choose aluminum foil. According to Nugroho & Redjeki (2015), the advantage of aluminum foil lies in its resistance to water vapor, gas, and physical and biological contamination. A'yun et al. (2023) explained that aluminum foil is more flexible, waterproof, airtight, heat-resistant, and non-toxic. From the overall consumer perception of the Dodol packaging redesigned with the Kansei Engineering approach, the packaging was very well received by respondents. All respondents tended to prefer the new packaging, which gives a premium impression. Based on data processing, there was a significant difference in respondents' perceptions of the old and new designs of 0.001 (<0.05), and the highest average value was found in the new secondary packaging design, namely 37.25. So, it can be concluded that in consumer perception, the new design is more attractive and increases purchasing demand compared to the old design. Emotional factors still play a role in consumer purchasing decisions. According to Koenig-Lewis et al (2014), emotions are the main drivers that influence decisions to purchase environmentally friendly products. This study evaluates environmental performance primarily from the perspective of consumer perception rather than through objective environmental assessment methods. Therefore, the findings should be interpreted as reflecting consumers' environmental aspirations and acceptance of packaging alternatives, rather than the actual environmental impacts of the materials. Future studies should integrate Life Cycle Assessment (LCA), carbon footprint analysis, or other quantitative sustainability assessment methods to provide a more comprehensive evaluation of packaging environmental performance.

ACKNOWLEDGMENTS

The author would like to thank the Jakarta State Polytechnic and the Packaging Printing Industrial Technology Study Program for their support in creating this research paper.

REFERENCES

A'yun, V. Q., Liviawaty, E., Pratama, R. I., & Junianto. (2023). Aluminium Foil as a Packaging

Material for Fishery Products: A Review. *Asian Journal of Fisheries and Aquatic Research*, *23*(3), 35–41. https://doi.org/10.9734/ajfar/2023/v23i3604

Ahmad, N. A. N., Lokman, A. M., Majid, E. S. A., & Suhaimi, A. I. H. (2025). The Kansei Engineering Method for Determining the Emotional Design Concept of the Metaverse Library Interface: Technical University Students Perspectives. *Journal of Technical Education and Training*, *17*(2), 53–67. https://doi.org/10.30880/jtet.2025.17.02.004

Atta-Delgado, M. X., Velazquez, G., Welti-Chanes, J., & Torres, J. A. (2024). Assessing consumer perceptions of sustainable packaging solutions. *CYTA - Journal of Food*, *22*(1). https://doi.org/10.1080/19476337.2024.2425403

Bou-Mitri, C., Abdessater, M., Zgheib, H., & Akiki, Z. (2021). Food packaging design and consumer perception of the product quality, safety, healthiness and preference. *Nutrition and Food Science*, *51*(1), 71–86. https://doi.org/10.1108/NFS-02-2020-0039

Chang, D., Lin, D., & Han, T. (2017). An ontology-based product affective properties identification approach. *Advances in Transdisciplinary Engineering*, *5*, 977–984. https://doi.org/10.3233/978-1-61499-779-5-977

Chukwurah, G. O., Okeke, F. O., Isimah, M. O., Enoguanbhor, E. C., Awe, F. C., Nnaemeka-Okeke, R. C., Guo, S., Nwafor, I. V., & Okeke, C. A. (2025). Cultural Influence of Local Food Heritage on Sustainable Development. *World*, *6*(1), 10. https://doi.org/10.3390/world6010010

Čok, V., Vlah, D., & Povh, J. (2022). Methodology for mapping form design elements with user preferences using Kansei engineering and VDI. *Journal of Engineering Design*, *33*(2), 144–170. https://doi.org/10.1080/09544828.2021.2012133

Eltohamy, M. S., Talaat, H. E. A., Moteleb, M. S. A., Mekhamer, S. F., & Omran, W. A. (2022). A Probabilistic Methodology for Estimating Reserve Requirement and Optimizing Its Components in Systems With High Wind Penetration. *IEEE Access*, *10*(October), 106148–106168. https://doi.org/10.1109/ACCESS.2022.3211305

Gao, P., Zhang, Y., & Long, Z. (2024). Kansei Drives Sustainable Material Innovation—An Approach to Enhance the Added Value of Biomass Materials. *Sustainability (Switzerland)* , *16*(13). https://doi.org/10.3390/su16135546

Hasan, B. M. S., & Abdulazeez, A. M. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *2*(1), 20–30. https://doi.org/10.30880/jscdm.2021.02.01.003

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202

Khan, H., & Lee, R. (2020). Does packaging influence taste and quality perceptions across varying consumer demographics? *Food Quality and Preference*, *84*(March), 103932. https://doi.org/10.1016/j.foodqual.2020.103932

Koenig-Lewis, N., Palmer, A., Dermody, J., & Urbye, A. (2014). Consumers' evaluations of ecological packaging - Rational and emotional approaches. *Journal of Environmental Psychology*, *37*, 94–105. https://doi.org/10.1016/j.jenvp.2013.11.009

Lin, Z. H., Woo, J. C., Luo, F., & Chen, Y. T. (2022). Research on Sound Imagery of Electric Shavers Based on Kansei Engineering and Multiple Artificial Neural Networks. *Applied Sciences (Switzerland)*, *12*(20). https://doi.org/10.3390/app122010329

Ma, L., Qing, C., Zhang, M., Zou, C., Pan, G., & Shen, Y. (2021). GWAS with a PCA uncovers candidate genes for accumulations of microelements in maize seedlings. *Physiologia Plantarum*, *172*(4), 2170–2180. https://doi.org/10.1111/ppl.13466

Mouludi, I., & Sipahutar, E. (2024). Tea Packaging Design Using Integration Of Kansei Engineering And Eye Tracking For Enhancing Consumer Appeal. *Journal of Engineering and Management in Industrial System*, *12*(2), 78–86.

Nagamachi, M. (2011). *Kansei/Affective Engineering*.

Nagamachi, M., & Lokman, A. (2015). *Kansei Innovation*.

Nugroho, A., & Redjeki, A. S. (2015). Pengaruh Waktu Pemanasan Pada Pembuatan Senyawa

Alum Dari Limbah Foil Blister Untuk Keperluan Industri Farmasi. *Jurnal Konversi*, *4*(2), 1. https://doi.org/10.24853/konversi.4.2.1-8

Otto, S., Strenger, M., Maier-Nöth, A., & Schmid, M. (2021). Food packaging and sustainability – Consumer perception vs. correlated scientific facts: A review. *Journal of Cleaner Production*, *298*. https://doi.org/10.1016/j.jclepro.2021.126733

Papantonopoulos, S., Bortziou, M., & Karasavova, M. (2021). A Kansei Engineering Study of Saffron Packaging Design. *International Journal of Affective Engineering*, *20*(4), 237–245. https://doi.org/10.5057/ijae.ijae-d-21-00006

Quispillo Parra, M. C. (2025). Estrategias tipográficas para reflejar la personalidad de una marca en distintos sectores industriales TT - Typographic Strategies for Reflecting a Brand's Personality Across Different Industrial Sectors. *Ñawi: Arte Diseño Comunicación*, *9*(1), 189–201.

Sari, N. P., Immanuel, J., & Cahyani, A. (2020). Aplikasi Kansei Engineering Dan Fuzzy Analytical Hierarchical Process Dalam Pengembangan Desain Kemasan. *Journal Printing and Packaging*, *1*, 9–21. http://jurnal.pnj.ac.id/index.php/ppt/article/view/2469/0

Sari, N. P., Prastiwinarti, W., Lokman, A. M., Yamin, I., Isna, A., Yennia, L., & Naufal, R. (2024). Developing the Concept of Emotion for Rendang Packaging Design Using Kansei Engineering. *Communications in Computer and Information Science*, *2313 CCIS*, 15–27. https://doi.org/10.1007/978-981-97-9890-2_2

Schütte, S., Lokman, A. M., Marco-almagro, L., Valverde, N., & Coleman, S. (2024). *Kansei for the Digital Era*. *23*(1), 1–18. https://doi.org/10.5057/ijae.IJAE-D-23-00003

Seref, N., & Cufaoglu, G. (2025). Food Packaging and Chemical Migration: A Food Safety Perspective. *Journal of Food Science*, *90*(5), 1–23. https://doi.org/10.1111/1750-3841.70265

Soenandi, I. A., Marcelle, M., Ondang, R. J., & Sundoro, A. N. (2021). *PERANCANGAN DAN PENGEMBANGAN PRODUK DESK ORGANIZER DENGAN METODE KANSEI ENGINEERING DAN MODEL KANO Iwan Aang Soenandi 1) , Marcelle 2) , Ronaldo Josua Ondang, Antika Ningrum Sundoro*. *9*(2), 117–128.

Sostar, M., Korov, T., & Pjero, E. (2025). The Impact of Packaging Type and Branding on Consumer Quality Perception and Preferences. *International Review of Management and Marketing* , *15*(5), 283–296. https://doi.org/10.32479/irmm.19336

Star Ria Sari, D., Alim Cendekia, D., & . S. (2020). The Aesthethic of Heritage Design of Siropen Premium Packaging in Holding Sustainability of Typical Souvenirs of Surabaya. *KnE Social Sciences*, *2020*, 145–161. https://doi.org/10.18502/kss.v4i12.7591

van Ooijen, I., Fransen, M. L., Verlegh, P. W. J., & Smit, E. G. (2017). Packaging design as an implicit communicator: Effects on product quality inferences in the presence of explicit quality cues. *Food Quality and Preference*, *62*, 71–79. https://doi.org/10.1016/j.foodqual.2017.06.007

Wibowo, M. A. S., Artini, W., Sidhi, E. Y., & Lisanty, N. (2025). Preferensi Konsumen Atas Pembaruan Kemasan Produk Kerupuk Lamuk (Studi Kasus Jajanan Tradisional Khas Desa Kalirong, Kediri). *Jurnal Agrica*, *18*(2). https://doi.org/10.31289/agrica.v18i2.14232

Yeoh, S. Y., Pua, Y., Tan, T. C., Seow, E. K., & Azhar, M. E. (2024). Impact of curdlan on quality and sensory attributes of canned dodol made from glutinous rice flour. *International Food Research Journal*, *31*(3), 601–613. https://doi.org/10.47836/ifrj.31.3.06

Yu, M., Abidin, S. Z., Shaari, N., He, C., Shi, L., & Liu, Q. (2024). Effects of brand visual identity on consumer attitude ： A systematic literature review. In *Environment and Social Psychology* (Vol. 9, Issue 9). https://doi.org/10.59429/esp.v9i9.3041

Zuo, Y., & Wang, Z. (2020). Subjective product evaluation system based on kansei engineering and analytic hierarchy process. *Symmetry*, *12*(8), 1–25. https://doi.org/10.3390/sym12081340

CLARIFICATION OF SOFTNESS PERCEPTION MECHANISM USING TETRA-PEG GEL

Keisuke NAKAJI[a], Aika YUKIMOTO[a] , Takeo KATO*[a] , Haruka ODA[b] , Yukitake NAKAHARA[a] , Makoto ASAI[a] , Takumi KATASHIMA[b]

[a] *Keio University, Japan*
[b] *The University of Tokyo, Japan*

ABSTRACT

This study tested the hypothesis regarding how physical interaction parameters - applied force, penetration, and contact area - between the object and the stiffer finger (4 out of 10 participants in this study had finger elastic moduli exceeding 50 kPa) influence softness perception. To this end, participants pressed Tetra-PEG gel with varying elastic moduli under two conditions: tool-based active touch and bare-based active touch, and during these interactions, applied force, penetration, and contact area were measured. The hypothesis was supported across nearly all participants.

Keywords: *softness, perception, active touch, deep sensation, cutaneous sensation*

1 INTRODUCTION

During the perception of object softness, research indicates that under bare-based active touch conditions, humans perceive softness using both cutaneous sensation and deep sensation, whereas under tool-based active touch conditions, they perceive softness using deep sensation alone (Lamotte & Srinivasan, 1995). Previous studies have shown that cutaneous sensation relies on contact area, while proprioception relies on force and penetration (Metzger et al., 2015; Scilingo et al., 2022; Bicchi A et al., 2000). Numerous existing studies on softness perception (Lamotte et al., 2000; Srinivasan & LaMotte, 1995; Bergmann, Tiest & Kappers, 2009; Xue et al., 2018) have measured one or two of the force, penetration, and contact area, with very few studies measuring all three simultaneously. Li et al. (2020, 2023) revealed that under bare-based passive touch conditions, significant differences between sample pairs emerge in the order of penetration, contact area, and force as the indentation depth increases. However, the relationship between force, penetration, contact area, and softness perception under bare-based active touch conditions remains unclear. During bare-based active touch, although cutaneous

* Corresponding author. Postal address: 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan. Email:kato@mech.keio.ac.jp

sensation serves as one of the cues, individuals with stiffer fingers have difficulty distinguishing softness based on contact area (Li et al., 2021), likely relying more on deep sensation. Therefore we hypothesize that participants with stiffer fingers perceive softness in a manner similar to conditions in which only deep sensation is available, even when cutaneous sensation is available. Specifically, we hypothesize that participants with stiffer fingers exhibit similar perceptual responses under both tool-based active touch and bare-based active touch conditions.

2 MATERIALS AND METHODS

Because object rigidity and elasticity influence softness perception (Bergmann, Tiest, & Kappers, 2009), this study utilized nine Tetra-PEG gels with equal rigidity but varying elasticity (32.9–64.6 kPa). Ten students (5 males, 5 females; mean age = 22.9 ± 0.83 years) participated. Their finger elastic moduli, measured following Han and Kawamura (1999), exhibited a bimodal distribution with a clear gap between 40.98 and 50.18 kPa (Table 1). Based on a 50 kPa threshold, four participants (Participants 3, 5, 7, and 9) were categorized into a higher stiffness group for the experiments.

During active touch, compressive force was measured using a semiconductor strain gauge. Penetration (maximum sample displacement) was recorded laterally via a camera (α9 (ILCE-9), Sony Corporation; Figure 1). Contact area was captured by a second camera viewing the finger-sample interface through a 45-degree mirror beneath the force plate; the pressure-darkened region (luminance $\leqslant$ 55 cd/m^2) was modeled as an ellipse to calculate this area.

For the procedure, participants first familiarized themselves with the samples by pressing all nine using tool-based active touch with their dominant index finger. Next, they alternately pressed paired samples five times each and selected the softer one, while force, penetration, and contact area were recorded. In total, 40 randomized trials were conducted (8 pairs of adjacent elastic moduli, tested five times each). This entire procedure was then repeated for bare-based active touch.

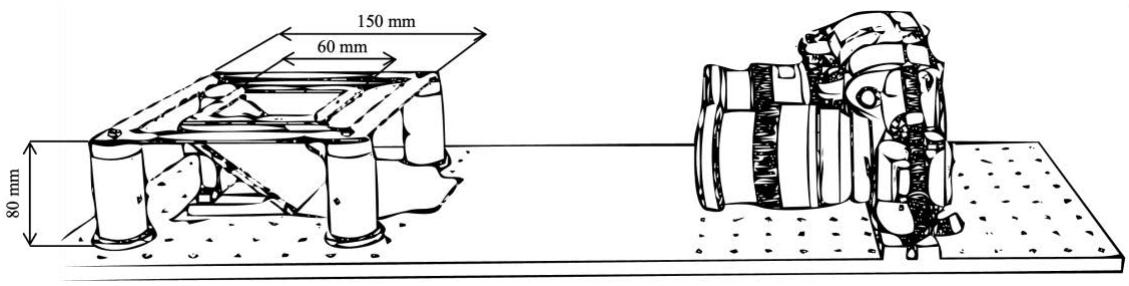


Figure 1. Apparatus for measuring penetration

Table 1. Size and Elastic Modulus of finger

Participant	1	2	3	4	5	6	7	8	9	10
Elastic modulus / kPa	38.82	40.98	62.97	37.57	56.92	38.22	50.18	35.68	68.91	33.19

3 RESULTS AND DISCUSSION

3.1 Result

First, we present the results for tool-based active touch. For each presented sample pair, a paired t-test (one-tailed) was performed by specifying the force and penetration measured when pressing high-modulus samples as the tF1 and tP1 groups, and those for low-modulus samples as the tF2 and tP2 groups. The alternative hypotheses were that the means of the tF1 and tP1 groups were greater than those of the tF2 and tP2 groups. Box-and-whisker plots showing the means, standard deviations (error bars), and t-test results for the tF1 vs. tF2 and tP1 vs. tP2 groups across all participants are presented in Figure 2. Here, * and ** denote significance levels of 0.05 and 0.01, respectively. This clearly shows that no significant difference is observed for force, whereas a significant difference is observed for penetration.

Next, regarding bare-based active touch, corresponding one-tailed t-tests were performed. The force, penetration, and contact area for high-modulus samples were designated as the bF1, bP1, and bC1 groups, and those for low-modulus samples as the bF2, bP2, and bC2 groups. The alternative hypotheses for force and penetration were the same as previously stated, while for contact area, the bC1 mean was hypothesized to be greater than the bC2 mean. Box-and-whisker plots showing the means, standard deviations (error bars), and t-test results for all participants in groups bF1 vs. bF2, bP1 vs. bP2, and bC1 vs. bC2 are presented in Figure 3. Significance levels are indicated as described earlier. Figure 3 reveals that for Participants 3, 5, and 9, no significant difference is observed for force, whereas a significant difference is observed for penetration;

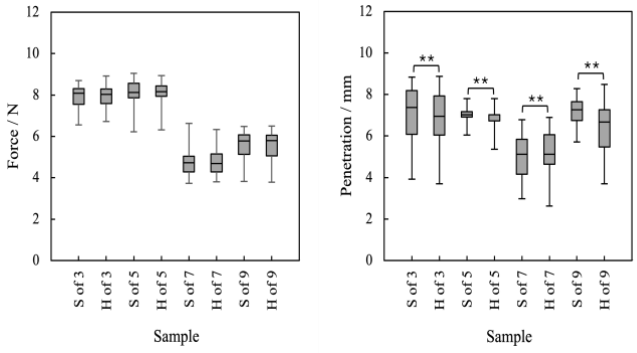


Figure 2. Force results (left) and penetration results (right)

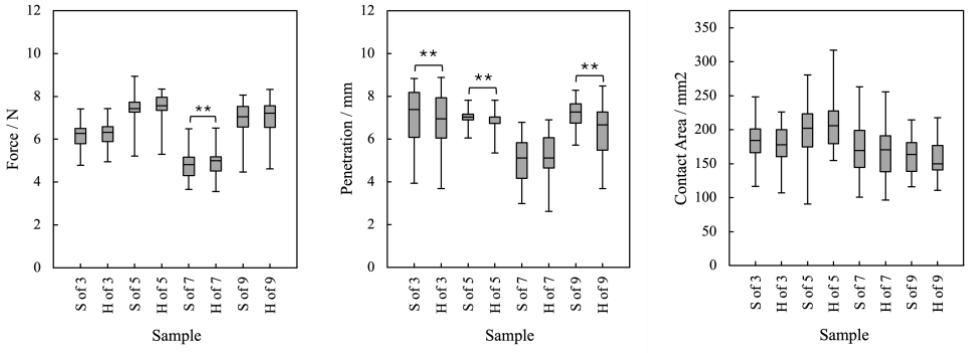


Figure 3. Force results (left), penetration results (center) and contact area results (right)

Participant 7 exhibits the opposite trend. Furthermore, no significant difference in contact area is observed for any participant.

To corroborate these findings, a logistic regression analysis was conducted on the experimental data. The results show that penetration has the largest partial regression coefficient (0.213) and demonstrates a marginally significant trend ($p = 0.083$).

3.2 Discussion

Participants 3, 5 and 9 evaluated the magnitude of penetration differences while maintaining similar compression forces under both conditions, yielding results that support Hypothesis. In contrast, participant 7 evaluated the magnitude of the difference in compression force while maintaining similar penetration, yielding results that do not support Hypothesis. One possible factor contributing to this is that, as revealed in the post-experiment questionnaire, this participant pushed the gel to the point where penetration no longer increased during bare-based active touch, and therefore likely thought they perceived the compression force at that stage. However, due to the limited number of participants in the high stiffness group (N=4), these findings should be considered preliminary observations. Future studies with a larger sample size are necessary to draw robust conclusions and further verify the present hypothesis.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI (grant number undetermined). We would like to thank the participants for their cooperation in the experiment.

REFERENCES

Bergmann Tiest, Wouter M. “Tactual Perception of Material Properties.” *Vision Research*, vol. 50, no. 24, Dec. 2010, pp. 2775–2782

Bicchi, A., et al. “Haptic Discrimination of Softness in Teleoperation: The Role of the Contact Area Spread Rate.” *IEEE Transactions on Robotics and Automation*, vol. 16, no. 5, 2000, pp. 496–504

LaMotte, Robert H. “Softness Discrimination with a Tool.” *Journal of Neurophysiology*, vol. 83, no. 4, 1 Apr. 2000, pp. 1777–1786

Li, Bingxu, et al. “Identifying 3-D Spatiotemporal Skin Deformation Cues Evoked in Interacting with Compliant Elastic Surfaces.” *PubMed Central*, 1 Mar. 2020, pp. 35–40,

Li, Bingxu, and Gregory J Gerling. “An Individual’s Skin Stiffness Predicts Their Tactile Discrimination of Compliance.” *The Journal of Physiology*, vol. 601, no. 24, 9 Nov. 2023, pp. 5777–5794

Li, Bingxu, and Gregory J. Gerling.. “Individual Differences Impacting Skin Deformation and Tactile Discrimination with Compliant Elastic Surfaces.” *2021 IEEE World Haptics Conference (WHC)*, 6 July 2021, pp. 721–726

Scilingo, Enzo Pasquale, et al. “Rendering Softness: Integration of Kinesthetic and Cutaneous Information in a Haptic Device.” *IEEE Transactions on Haptics*, vol. 3, no. 2, Apr. 2010, pp. 109–118

Srinivasan, Mandayam A, and Robert H LaMotte. “Tactual Discrimination of Softness.” *Journal of Neurophysiology*, vol. 73, no. 1, 1 Jan. 1995, pp. 88–101,

Xue Song Zhou, et al.“*Correlation between Tactile Perception and Tribological and Dynamical Properties for Human Finger under Different Sliding Speeds”*. Vol. 123, 1 July 2018, pp. 286–295

KEER2026, EUGENE, OR, USA | AUGUST 11-12, 2026
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2026

COMPANIONSHIP LIGHTING

A MULTI-SENSORY COMPANION LIGHTING SYSTEM FOR SOCIAL CONNECTION

Muthita Wongngan[1], Wang Tzu Jung[2]
[1]Department of Arts and Design, National Tsing Hua University, Hsinchu, Taiwan
[2]Department of Arts and Design, National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT

This study addresses the critical lack of emotional support and the persistent need for non-intrusive companionship in modern society. The design challenge is framed by the observation that fear of rejection creates a vicious cycle where individuals avoid initiating social contact. The research seeks to develop a low-burden mechanism that lowers the psychological threshold for seeking companionship, moving beyond the functional limitations of typical digital interactions.

The research employed a mixed-methods approach, combining quantitative Kansei Engineering Type II analysis with qualitative study. The Semantic Differential (SD) method was used to assess the emotional attributes of 15 existing lamp designs. Qualitative analysis of five high-performing samples (S3, S6, S7, S8, S11) deconstructed material and formal strategies. Both results informed the development of a prototype: a warm-toned LED lighting fixture with a mobile application offering touch-to-signal online status and conditional one-touch invitations for shared activities.

Principal Component Analysis identified three factors explaining 90.85% of variance: Natural Warm Atmosphere (68.42%), Comfort (16.00%), and Ceremony and Quality (6.43%). Critically, Have Companion–Lonely loaded strongly (0.867) within Natural Warm Atmosphere, revealing that companionship perception correlates with material naturalness rather than technological features. Qualitative analysis revealed existing companion lamps face a fundamental trade-off: designs maximizing natural warmth (soft materials, compact scale ≤7cm) lack ceremonial gravitas, while ceremonial designs feel commercial. The prototype occupies this strategic gap through handmade paper clay base (≤7cm) for tactile approachability with hand-blown frosted glass for perceived craftsmanship. This research offers three actionable principles: (1) material naturalness may be more important than technological interactivity for companionship perception, (2) physical constraint enables ritual integration when paired with premium material quality, and (3) ambiguous availability cues eliminate social friction while preserving user autonomy.

Keywords: Emotional Design, Kansei Engineering, Multi-sensory Interaction, Companionship

1 INTRODUCTION

According to the World Health Organization, approximately one in six people globally is currently affected by loneliness, a phenomenon that has significantly impacted physical and mental health. Although digital communication technologies have achieved unprecedented

density in information connection, individual loneliness has not decreased; instead, it presents the paradox of being "Alone Together" (Turkle, 2011). Existing Instant Messaging (IM) and video conferencing tools, while efficient, often impose high cognitive loads and social anxiety. This "always-on" mode of interaction is frequently perceived as an intrusive disturbance, failing to convey non-verbal companionship.

To address these challenges, this study proposes "SyncLight," a hybrid system integrating an interactive lamp with a low-burden social application. The design aims to balance personal solitude with social connection by allowing lighting behavior to signal availability without verbal intrusion. This paper details the system's architecture and investigates its perceived value as a ritual-oriented mechanism through qualitative interviews and prototype testing.

2 LITERATURE REVIEW

2.1 Remote Companionship and Ambient Mediation Technologies

The paradigm of remote interaction has evolved from functional information transmission toward the cultivation of affective relatedness. Hassenzahl et al. (2012) posited that contemporary technology design must transcend data exchange to mediate psychological proximity. Brave et al. (1998) introduced synchronized distributed physical objects, wherein virtual signals are materialized into tactile forms. Rieger (2019) demonstrated through LumiTouch how tactile gestures translate into ambient light patterns, enabling sustained emotional bonding under low-effort conditions.

2.2 Social Initiation Friction and Low-Burden Interaction Design

Despite ubiquitous connectivity, the psychological friction of social initiation persists. Turkle (2011) articulated the paradox of "alone together," wherein individuals gravitate toward controlled companionship that minimizes social vulnerability. This intensifies among populations with elevated fear of negative evaluation (Iqbal, 2019). Birnholtz et al. (2010) argued that effective design must incorporate ambiguous availability cues—signals conveying presence without demanding immediate response.

2.3 Ritual Design and Temporal Engagement

Ozenc and Hagan (2018) demonstrated that ritual design—structured interactive behaviors with symbolic meaning—effectively regulates affect. Operationalizing technological engagement as daily ritual confers biomimetic vitality upon objects, reframing digital connectivity as long-term symbiotic relationships.

3 METHODS

This study employs a Mixed-Methods Research Design, combining quantitative Kansei Engineering Type II (KE-II) with qualitative analysis. The study follows a four-phase framework: (I) Product Analysis, (II) Quantitative Kansei Engineering Analysis, (III) Prototype Implementation, and (IV) Qualitative Validation.

3.1 Phase I: Product Analysis

A systematic collection of 30 existing lamp designs was conducted using keywords such as 'aroma lamp,' 'mood lighting,' and 'therapeutic lighting.' Through selection criteria ensuring high diversity, 15 representative lamp designs (S1 to S15) were selected as stimuli.

Bipolar adjective pairs were systematically compiled from three research domains: lighting semiotics (Albertazzi, 2018; Castilla et al., 2018), olfactory affective space (Delplanque et al., 2012; Ferdenzi et al., 2013), and cross-modal environmental perception (Zuo et al., 2023). An initial pool of 20 pairs was reduced to 10 through pilot testing with 5 participants, selecting pairs with highest discriminatory power and lowest inter-correlation ($r < 0.70$).

The final set comprised four perceptual dimensions: Color Temperature (Warm–Cold, Soft–Dazzling), Emotional (Safe–Anxious, Relaxing–Stimulating), Context (Homely–Commercial, Natural–Artificial, Have Companion–Lonely), and Overall Experience (Comfortable–Uncomfortable, Ceremonial–Non-Ceremonial, Cheap–Expensive).



Figure 1. *Sample items representing three clusters (Natural, Metal Technology, Mixed)*

3.2 Phase II: Quantitative Kansei Engineering Analysis

The target audience comprised adults aged 25-34 years (70%) and 18-24 years (26.6%), aligning with urban workers experiencing high stress and loneliness. Thirty participants rated the 15 lamp images based on 10 bipolar adjective pairs using a seven-point Semantic Differential scale.

The unit of analysis was the lamp sample: for each of the 15 samples, ratings were averaged across the 30 participants on each adjective pair, yielding a 15 (samples) × 10 (adjective pairs) mean-score matrix. Principal Component Analysis (PCA) was then performed in SPSS on the correlation matrix of the 10 adjective pairs, applying Varimax rotation with Kaiser normalization to reduce dimensionality.

3.3 Phase III: Prototype Implementation

Based on quantitative findings, the SyncLight prototype integrates hardware and software to occupy the identified market gap.

Adjective Pair	Factor 1	Factor 2	Factor 3

Natural–Artificial	.950	.001	-.231
Homely–Commercial	.924	.266	.033
Have Companion–Lonely	.867	.159	.115
Comfortable–Uncomfortable	.294	.936	.165
Ceremonial–Non-Ceremonial	.211	.199	.881

Table 1. *Rotated Component Matrix (Selected Loadings >0.80)*

Factor 1 is dominated by Natural (.950), Homely (.924), and Have Companion (.867). Factor 2 is defined by Comfortable (.936). Factor 3 is characterized by Ceremonial (.881).

Sample	Image	Factor 1	Factor 2	Factor 3	Material
S6		+1.43	+0.98	-0.76	Soft felt
S7		+1.69	+1.69	-0.38	Concrete + amber glass
S8		+1.01	+0.58	-0.73	Stone-like translucent
S11		+1.58	-0.53	+1.36	Embossed paper
S3		-0.63	-0.14	+1.65	Faceted glass

Table 2. *Sample Coding Analysis showing factor scores for five high-performing samples*

Qualitative analysis of five high-performing samples revealed material strategies: S6 and S7 maximized natural warmth and comfort through soft materials, while S3 and S11 achieved high ceremonial quality through premium materials. This exposed a strategic gap: no existing product simultaneously achieves high scores across all three factors.



Figure 2. *Prototype combines stone-like base (paper clay) for natural warmth + hand-blown frosted glass for premium craftsmanship*

Hardware Design. The prototype employs handmade paper clay base (maximum 7cm height) mimicking natural stone texture, addressing Factor 1 attributes (Natural: 0.950, Homely: 0.924, Safe: 0.883). Hand-blown frosted glass diffusers provide soft light distribution (Soft: 0.865, Comfortable: 0.936). Integrated scent diffusion enhances multi-sensory relaxation. Warm-toned LED (2700-3000K) with breathing rhythm supports emotional regulation.

Software Design. The SyncLight mobile application comprises four modules: (1) Touch-to-Online Status Synchronization - tapping the lamp updates online status visible to friends; (2) Conditional One-Touch Invitation - activity invitations appear only when both parties are mutually online, eliminating rejection fear; (3) Companion Growth System - gamifies engagement through digital pet metaphor where the lamp grows with interaction frequency; (4) Shared Time Visualization - displays companionship through rankings and usage calendar.



Figure 3. *App screens showing (left) online status synchronization, (center) companion rankings, (right) usage calendar*

3.4 Phase IV: Qualitative User Validation

Semi-structured interviews with 7 participants followed a dual-phase process: (1) Blind Test—intuitive perception without functional explanation; (2) Contextual Inquiry—re-evaluation after understanding design logic. Interview procedure included: product demonstration, tactile interaction (10-20 seconds), feature introduction, semi-structured interview, and design feedback. Analysis employed thematic coding and affinity diagramming.



Figure 4. *(left) Tactile interaction during blind test, (right) Affinity diagramming process*

4 RESULTS

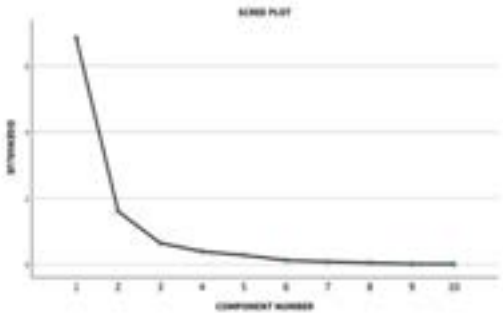
4.1 Kansei Engineering Findings

Three components were retained based on the scree-plot elbow (Figure 5) and interpretability, the third capturing a theoretically important Ceremony/Quality dimension despite an eigenvalue slightly below 1.0. The three components collectively explain 90.85% of total variance.

Component	Total	% Variance	Cumulative %	% Variance	Cumulative %
1	6.842	68.417	68.417	68.417	68.417
2	1.600	16.002	84.420	16.002	84.420
3	0.643	6.428	90.847	6.428	90.847

Table 3. *Total Variance Explained by Principal Component Analysis*

Factor 1 (Natural Warm Atmosphere) accounts for 68.42% of variance with an eigenvalue of 6.842. Factor 2 (Comfort) contributes 16.00% (cumulative 84.42%), and Factor 3 (Ceremony and Quality) adds 6.43% (cumulative 90.85%).



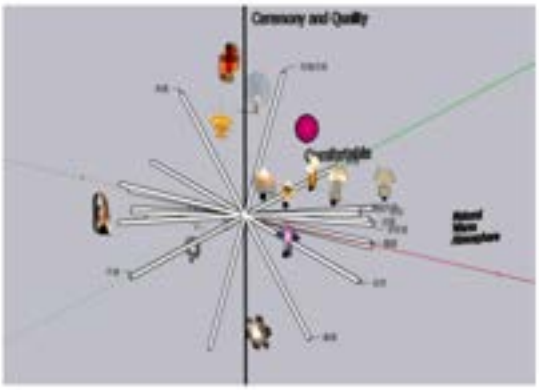


Figure 5. *Scree plot with clear elbow at component 3* ***Figure 6.*** *Positioning map (pink circle)*

4.2 Strategic Positioning Gap

Multi-dimensional scaling visualized the positioning of 15 lamp samples within the three-factor space. The analysis revealed a unique market gap (pink circle in Figure 6) combining high Natural Warm Atmosphere, Comfort, and Ceremony/Quality—largely unoccupied by existing products. High-natural lamps (felt, paper materials; ≤7cm) lacked ceremonial quality. Ceremonial lamps (glass, metal; complex forms) felt commercial rather than homely.

4.3 User Interview Themes

Thematic analysis identified four major themes:

Theme 1: Materiality of Interaction. Natural, tactile materials create emotional grounding and therapeutic value. Participants valued materials providing both tactile comfort and symbolic connection to nature. Warm color temperature (2700K) was universally preferred; breathing pattern creates "living presence." Scent was valued as optional enhancement, not core feature.

Theme 2: Social Connection Design. Critical finding: dual-mode functionality essential—some users require solitary operation without social pressure. User 7 emphasized: "Product should work without app dependency... Some need to be comfortable alone first." Conditional invitation mechanism effectively reduces rejection anxiety. Users demand high degree of control over social availability and growth features.

Theme 3: Form and Aesthetic Preferences. Small, rounded forms perceived as friendly and approachable, aligning with research on curved visual preference (Bar & Neta, 2006). The compact scale (≤7cm) was identified as optimal for bedside placement. Handmade, unique bases significantly increase emotional attachment through the IKEA effect (Norton et al., 2012).

Theme 4: Engagement and Retention. Critical problem identified: users with no existing friends on platform lack engagement drivers. Proposed solutions include workshop onboarding (participants automatically connected) and standalone value (lamp remains functional without social features). User 1 noted: "Options within options" allowing users to modulate engagement level based on life circumstances.



Figure 7. *Typical evening usage scenarios illustrating SyncLight in bedroom contexts*

5 DISCUSSION

This research makes three key contributions to companion-oriented design:

Material Naturalness May Be More Important than Technological Interactivity. The finding that "Have Companion–Lonely" loaded strongly (0.867) within Natural Warm Atmosphere—rather than forming an independent factor—suggests that digital interactivity may not be the primary driver of companionship perception. Since technological interactivity was not a directly measured variable in the quantitative study, this interpretation should be treated as tentative; nevertheless, qualitative analysis confirmed users seek objects that "co-exist" rather than "respond." This suggests designers should invest in tactile craftsmanship before adding smart features.

Physical Constraint as Ritualization Asset. Competitive analysis revealed high-natural lamps were constrained to soft materials and ≤7cm height, limiting ceremonial perception. The SyncLight prototype reframes compact scale as enabling intimate ritual integration. Pairing small dimensions with premium materials (hand-blown glass) resolves the naturalness-ceremony trade-off. The ≤7cm dimension positions the lamp as a bedside companion—small enough to not dominate space yet substantive enough to command daily attention.

Ambiguous Availability Cues for Social Friction Reduction. The conditional mutual-visibility mechanism transforms social initiation from explicit request (vulnerable to rejection) into background presence awareness. Unlike traditional IM systems where users must actively message and risk rejection, SyncLight's touch-to-online mechanism broadcasts availability passively. Activity invitations appearing only when both parties signal presence eliminates fear of "bothering someone" while preserving user autonomy.

Limitations. This study has several limitations. First, the sample size for qualitative validation (7 participants) is relatively small. Longitudinal studies tracking sustained engagement over months are necessary to validate the companion growth system's effectiveness. Second, the study focused on Taiwanese young adults; cross-cultural validation is needed. Third, the cold-start problem requires practical solutions before commercial deployment—workshop-based onboarding or social seeding strategies warrant further investigation.

6 CONCLUSION

This study addressed the challenge of designing companion technologies that reduce loneliness without imposing social burden. Through mixed-methods analysis, we identified a strategic market gap where no existing product simultaneously achieves natural warmth,

comfort, and ceremonial quality. The SyncLight prototype occupies this position through material strategy (handmade paper clay + hand-blown glass) and interaction design (conditional mutual-visibility + one-touch invitations).

Key findings challenge prevailing assumptions: (1) companionship perception correlates strongly with material naturalness (0.867), suggesting that material naturalness may be more important than interactive complexity; (2) physical constraint (≤7cm) enables ritual integration when paired with premium materials; (3) ambiguous availability cues eliminate social friction while preserving autonomy. These principles offer designers a replicable framework for creating emotionally resonant companion artifacts.

Three design principles emerge: Natural Materials and Tactile Qualities—designers should select materials with warm, natural surfaces (wood, stone, ceramic, fabric) providing psychological safety. Multi-Sensory Interaction—combining visual (warm light), tactile (natural surfaces), and olfactory (aromatherapy) experiences enhances relaxation. Light Design—warm-toned light (2700-3000K) and slow breathing rhythms (12+ seconds/cycle) support emotional grounding. Supplementary principle: Design for Autonomy—products should function without app dependency; all connectivity and gamification features should be toggleable.

As loneliness escalates as a global health crisis, the design community bears responsibility for creating technologies fostering genuine connection. This research contributes actionable design principles grounded in empirical evidence, offering a path toward companion technologies that truly serve human emotional needs.

REFERENCES

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. Psychological Science, 17(8), 645-648.

Birnholtz, J., Lai, Y., Shih, L., & Munson, S. (2010). Social and communicative aspects of always-on presence. In Proceedings of CHI 2010 (pp. 2885-2888). ACM.

Brave, S., Ishii, H., & Dahley, A. (1998). Tangible interfaces for remote collaboration. In Proceedings of CSCW 1998 (pp. 169-178). ACM.

Castilla, N., Llinares, C., Bisegna, F., & Blanca-Giménez, V. (2018). Emotional evaluation of lighting in university classrooms. Frontiers of Architectural Research, 7(4), 600-609.

Delplanque, S., et al. (2012). Emotional processing of odors. Chemical Senses, 37(5), 469-479.

Ferdenzi, C., et al. (2013). Affective semantic space of scents. Food Quality and Preference, 30(2), 128-138.

Hassenzahl, M., et al. (2012). All you need is love: Mediating intimate relationships through technology. ACM TOCHI, 19(4), Article 30.

Iqbal, A. (2019). Fear of negative evaluation and social anxiety in young adults. Peshawar Journal of Psychology and Behavioral Sciences, 5(1), 63-76.

Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces. In Proceedings of CHI 1997 (pp. 234-241). ACM.

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books.

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. Journal of Consumer Psychology, 22(3), 453-460.

Ozenc, F., & Hagan, M. (2018). Ritual design: Crafting team rituals for meaningful organizational change. In Service Design Geographies (pp. 146-157).

Rieger, S. (2020). Reduction and participation. In The democratization of artificial intelligence (pp. 143-162). transcript Verlag.

Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other. Basic Books.

Weiser, M., & Brown, J. S. (1995). Designing calm technology. PowerGrid Journal, 1(1), 75-85.

Zuo, T., et al. (2023). The impact of indoor environmental quality on occupant satisfaction. Building and Environment.

Cross-Cultural Analysis of Design Elements for Gender-Neutral Packaging: A Kansei Engineering Study Comparing Thailand and Taiwan

Huang Ying Jhen[1], Muthita Wongngan[2], Tsuei Ju Hsieh[3]

[1]Department of Arts and Design, National Tsing Hua University, Hsinchu, Taiwan
[2]Department of Arts and Design, National Tsing Hua University, Hsinchu, Taiwan
[3]Department of Arts and Design, National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT

The rapid growth of gender-neutral design in fragrance, cosmetics, and personal care markets has created an urgent need for empirical design frameworks. This study applies Kansei Engineering methodology (Nagamachi, 2002) to investigate how packaging design elements—color, typography, and surface finish—influence gender-neutral consumer perception across two culturally distinct samples in Thailand and Taiwan. Twelve packaging stimuli spanning fragrance, cosmetics, and personal care categories were evaluated using a four-item Semantic Differential Scale derived from the brand-gender literature (Grohmann, 2009; Lieven et al., 2015): Delicate–Strong, Feminine–Masculine, Static–Dynamic, and Decorative–Minimalist. Multidimensional scaling (MDS) with Ward hierarchical cluster analysis mapped the perceptual proximity of the twelve stimuli in each market, revealing culturally distinct cluster structures: grayscale-based designs occupied the neutral core in both markets, whilst earth-tone designs shifted from neutral-organic (Thailand) toward masculine-coded (Taiwan). Quantification Theory Type I determined category scores and importance weights. Primary Color was the dominant element in both samples (Thailand: 42.8%; Taiwan: 32.6%), with grayscale and earth tones achieving the highest Gender-Neutrality Scores. Cross-cultural divergences emerged: within the dominant colour element, earth tones achieved the highest neutrality in Thailand whilst grayscale ranked highest in Taiwan; Font Type was more salient in Thailand (28.2%) than Taiwan (20.4%); Surface Finish more salient in Taiwan (25.5%) than Thailand (16.1%); and Font Weight showed notable divergence (Taiwan: 21.5% vs. Thailand: 12.9%). A 'clinical-minimalist' design archetype—neutral palette, sans-serif typography, matte surface—achieved cross-cultural consensus as the most reliable gender-neutral configuration. A generational subgroup analysis further showed that Gen Z and Gen Y rankings converge in Taiwan (rs = .95) but diverge in Thailand (rs = .60), and that surface-finish salience is a Generation Z characteristic in both countries. Three design prototypes operationalise findings as concrete packaging concepts for practitioners.

Keywords: *gender-neutral packaging, Kansei Engineering, cross-cultural design, Quantification Theory Type I, inclusive design*

1. INTRODUCTION

1.1 Research Background and Motivation

The global fragrance, cosmetics, and personal care market is experiencing a fundamental transformation in brand identity strategy. A three-phase evolution is observable: pre-2000 packaging relied on explicit gender categorisation through pink-for-female and navy-for-male conventions; 2000–2015 introduced unisex positioning with reduced gender divisions; and since 2015, intentional gender-neutral design has emerged as a dominant

premium strategy, exemplified by Aesop, Byredo, Le Labo, and CK One. Market research indicates that 67% of Generation Z consumers prefer products without explicit gender definitions (Mintel, 2022), and the global gender-neutral cosmetics market is projected to grow at a compound annual growth rate of 8.5% through 2030 (Grand View Research, 2023).

1.2 Research Gap and Objectives

Grohmann's (2009) seminal Brand Personality Gender Framework established that Masculine Brand Personality (MBP) and Feminine Brand Personality (FBP) function as independent dimensions, and identified 'undifferentiated' brands (low MBP, low FBP) as warranting further study. Grohmann called for 'future research to examine the effects of gender perceptions on consumer responses more thoroughly... particularly for undifferentiated brands, which have received little attention in prior research' (p. 116). This study addresses that gap through Kansei Engineering applied to cross-cultural samples in Thailand and Taiwan, identifying which specific packaging design elements—and at what relative importance—produce gender-neutral perception in each market, and which configurations remain universally neutral across both markets and consumer generations.

2. LITERATURE REVIEW

2.1 Brand Personality Gender Dimensions

Grohmann (2009) established that brand masculinity and femininity function as two independent dimensions, positioning brands in four quadrants: Masculine (high MBP, low FBP: Nike, Marlboro), Feminine (low MBP, high FBP: Chanel, Dove), Androgynous (high both), and Undifferentiated (low both: Aquafina, Benadryl). Contemporary gender-neutral premium brands (Aesop, CK One) occupy the Undifferentiated quadrant but differ from Grohmann's examples through intentional premium positioning and inclusive brand values rather than accidental identity deficit.

2.2 Kansei Engineering and Quantification Theory Type I

Kansei Engineering (Nagamachi, 2002) links consumer psychological responses to concrete design parameters through a five-step process: Kansei word collection, semantic differential measurement, statistical analysis, design rule derivation, and validation. Quantification Theory Type I (Hayashi, 1952), the analytical centrepiece of this study, extends multiple regression to categorical independent variables (colors, fonts, materials), producing category scores and importance weights that translate directly into design decision tools. The mathematical formulation $Y = \beta_0 + \Sigma_i \Sigma_j \beta_{ij} \delta_{ij}$ enables prediction of Gender-Neutrality Scores from design element combinations prior to physical prototyping.

2.3 Visual Design Elements and Gender Perception

Van Esch et al. (2020) established color as the most immediate and potent gender-coding cue in packaging, processed pre-attentively before any conscious evaluation. Orth and Malkewitz (2008) demonstrated that packaging creates holistic impressions through the synergistic combination of elements—no single element can be evaluated in isolation. Norman's (2004) Emotional Design Theory further establishes that visceral-level elements (color, typography, texture) must align with reflective-level meaning (inclusive values, quality positioning) for coherent gender-neutral brand communication.

2.4 Theoretical Basis for Kansei Word Selection

The four Kansei word pairs used in this study were selected on theoretical and empirical grounds. The Feminine–Masculine dimension derives directly from Grohmann's (2009) Brand Personality Gender Scale, which established masculinity and femininity as the two primary axes of brand gender perception. The Delicate–Strong dimension reflects Lieven et al.'s (2014, 2015) finding that perceived power or delicacy is the second major component of

brand gender evaluation, independently validated across multiple product categories. The Decorative–Minimalist dimension captures visual complexity, identified by Orth and Malkewitz (2008) as a central holistic dimension of packaging perception and confirmed as a universal design evaluation schema across cultures. The Static–Dynamic dimension reflects the energy or vitality perceived from packaging form and typography, a dimension identified in Henderson et al.'s (2004) work on typeface personality and confirmed through pilot testing to be orthogonal to gender coding in Thai consumer samples but partially gender-associated in Taiwanese samples. Candidate word pairs were initially drawn from these four theoretical sources, reduced through expert review (N = 2 experienced design researchers) and a pilot test (N = 5 designers), retaining the single most representative pair per dimension based on internal consistency (pilot r = .884–.912).

3. METHODOLOGY

3.1 Sample Collection and Expert Validation

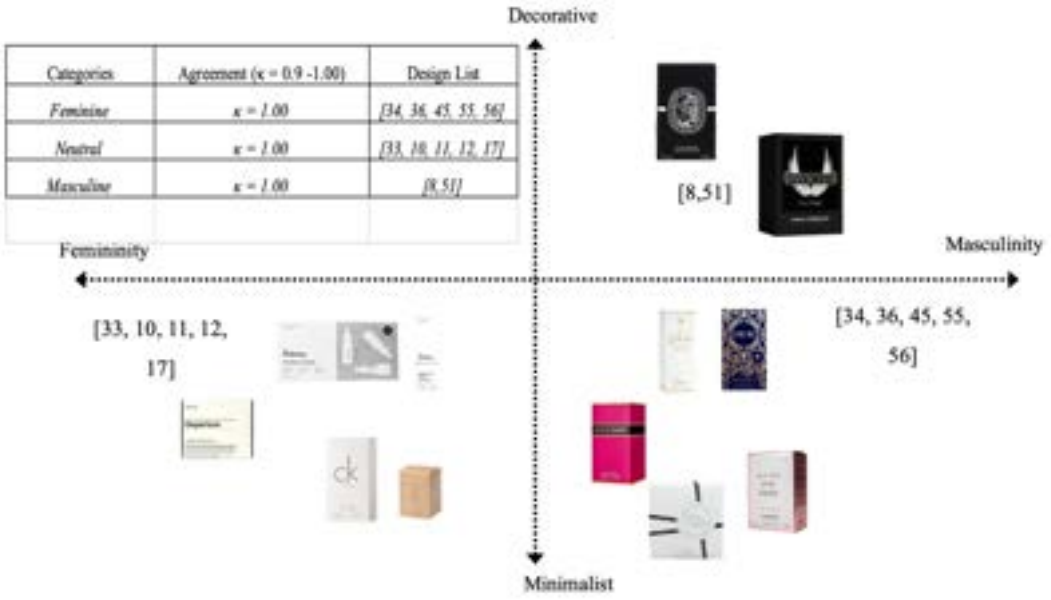
Sixty premium packaging designs were initially collected from fragrance, cosmetics, and personal care categories. Five expert designers with seven or more years of professional experience independently categorised each design as Feminine, Neutral, or Masculine without consultation, preserving independence of judgement. Only designs achieving perfect inter-rater consensus (Fleiss' $\kappa = 1.00$) were retained for consumer evaluation, yielding twelve validated stimuli: five Feminine, five Neutral, and two Masculine. All twelve stimuli received consumer survey evaluation and are detailed in Table 1.

Table 1. Validated packaging stimulus set: design element coding and GNS (N = 12)

ID	Design	Picture	Font Weight	Surface	Cat.	GNS (TH)	GNS (TW)
P1	Prada Candy		Sans-serif Regular	Glossy	Fem.	7.705	6.593
P2	Diptyque Eau Capitale		Serif Regular	Semi-gloss	Neu.	6.744	6.637
P3	The Ordinary Niacinamide		Sans-serif Bold	Matte	Neu.	7.538	7.456
P4	Lancôme		Sans-serif Light	Semi-gloss	Fem.	7.532	7.127
P5	The Ordinary Set Box		Sans-serif Bold	Matte	Neu.	7.833	7.387
P6	Dior J'adore		Script Bold	Matte	Fem.	7.109	7.186

P7	Dior Sauvage		Sans-serif Regular	Glossy	Masc.	6.795	6.765
P8	MUJI Container		Sans-serif Regular	Matte	Neu.	7.929	6.912
P9	Viktor & Rolf		Sans-serif Regular	Semi-gloss	Neu.	7.840	7.108
P10	Aesop Departure		Sans-serif Bold	Matte	Neu.	7.872	7.250
P11	CK One (Calvin Klein)		Sans-serif Light	Matte	Neu.	7.583	7.549
P12	Paco Rabanne Invictus		Script Light	Semi-gloss	Masc.	6.445	6.728
	Sample mean GNS			7.410	7.058		

Note. Cat. = expert-validated category; GNS = Gender-Neutrality Score (max. 8.0). Neu. = Neutral, Fem. = Feminine, Masc. = Masculine.



3.2 Kansei Word Selection and Semantic Differential Instrument

The four Kansei word pairs were selected through a literature-driven funnel followed by empirical validation, as described in Section 2.4. Candidate pairs were drawn from Grohmann (2009), Lieven et al. (2014, 2015), Orth and Malkewitz (2008), and Henderson et al. (2004), then refined through expert review (N = 2) and pilot testing (N = 5). The final four dimensions—Delicate–Strong, Feminine–Masculine, Static–Dynamic, and Decorative–Minimalist—each demonstrated strong pilot reliability (r = .884–.912). Each pair was operationalised as a 7-point Semantic Differential Scale (1 = left pole, 7 = right pole; 3.5–4.5 = neutral zone).

3.3 Consumer Perception Survey

The survey was administered via Qualtrics to two national samples targeting consumers aged 16–45 who regularly purchase fragrance, cosmetics, or personal care products. The Thailand sample comprised n = 78 respondents and the Taiwan sample n = 102 respondents, for a combined total of N = 180. Both samples were predominantly Generation Y and Z, educated to bachelor's degree level. The Thai sample was notably more diverse in gender identity, with 40 respondents (51.3%) identifying as non-binary or they/them. Table 2 presents the full demographic composition. The survey presented all twelve packaging stimuli in randomised order. Each respondent rated all twelve designs on the four Semantic Differential scales, producing 8,640 individual ratings (180 × 12 × 4).

Table 2. Sample composition by country

Demographic	Thailand (n = 78)	Taiwan (n = 102)
Gender: Female	23	71
Gender: Male	11	30
Gender: Non-binary / They-Them	40	1
Gender: Prefer not to say	4	0
Age: Gen Z (≤ 30)	41	56
Age: Gen Y / older (31+)	37	46

Note. Survey administered in Thai, English, and Chinese.

3.4 Gender-Neutrality Score (GNS)

A two-dimensional Gender-Neutrality Score was operationalised consistent with Grohmann's (2009) MBP/FBP framework, incorporating both the Feminine–Masculine dimension (gender identity) and the Delicate–Strong dimension (perceived power), the two components of brand gender perception established by Grohmann (2009) and Lieven et al. (2015):

$$\mathbf{GNS = 8 - (|M(Delicate–Strong) - 4| + |M(Feminine–Masculine) - 4|) \div 2}$$

A GNS of 8.00 represents perfect gender neutrality (both dimensions at midpoint 4.00); lower values indicate increasing gendered perception. This two-dimensional operationalisation was preferred over single-dimension measures because brand gender perception comprises both identity and power components (Grohmann, 2009; Lieven et al., 2015), and the two dimensions were strongly correlated in both national samples.

4. RESULTS

4.1 Stimulus Set and Descriptive Statistics

Design element coding and GNS values for the twelve validated stimuli are reported in Table 1 (Section 3.1). Table 3 reports mean Kansei dimension scores.

Table 3. Descriptive statistics of Kansei evaluation scores: Thailand and Taiwan

Kansei Word Pair	TH M	TH SD	TW M	TW SD	Scale
Delicate–Strong	3.998	1.849	4.426	1.415	1–7
Feminine–Masculine	3.955	1.442	3.681	1.344	1–7
Static–Dynamic	4.521	1.229	3.848	1.304	1–7
Decorative–Minimalist	3.699	1.291	4.523	1.155	1–7

N (observations)	936	—	1,224	—	—

Note. Each observation = one respondent × one stimulus × one scale item.

4.2 Multidimensional Image Space: Perceptual Maps and Cluster Structure

Following the analysis plan, multidimensional scaling (MDS) with Ward hierarchical cluster analysis (HCA) was applied to visualise the perceptual proximity of all twelve stimuli within each market. Figures 2 and 3 present the MDS perceptual maps with k = 3 cluster solutions; Figures 4 and 5 present the corresponding dendrograms.

In Thailand (Figure 2, Figure 4), the three-cluster solution separates: (1) a gendered-masculine cluster (P2 Diptyque, P7 Dior Sauvage, P12 Invictus) clearly isolated on the right of the map; (2) a compact minimalist-neutral pair (P3 The Ordinary Niacinamide, P8 MUJI) in the lower-left; and (3) a large central cluster containing the remaining feminine and neutral designs. The Thai perceptual space thus separates designs primarily along the gender axis, with the masculine trio at maximal Ward distance (≈ 7) from all other stimuli.

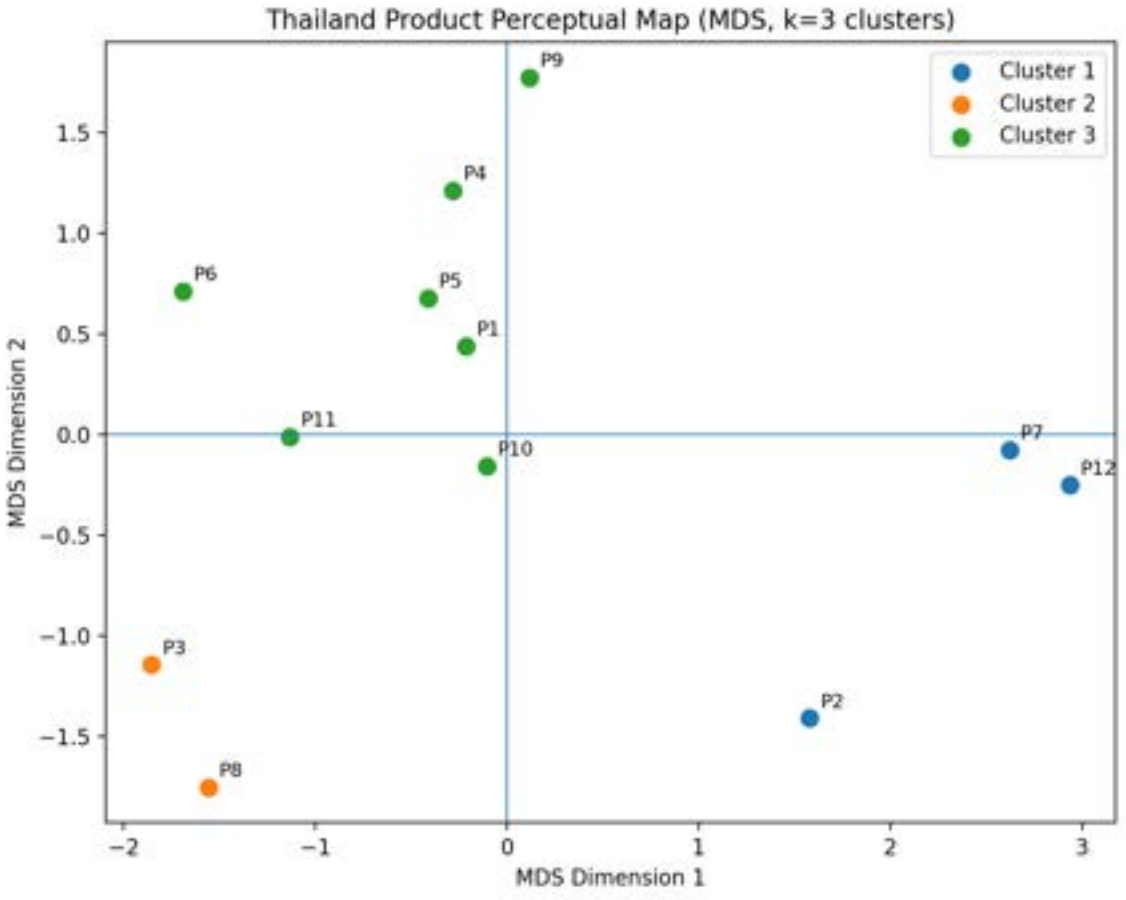


Figure 2. Thailand product perceptual map (MDS, k = 3 clusters). Cluster 1 (blue) = masculine-coded designs; Cluster 2 (orange) = minimalist-neutral exemplars (P3, P8); Cluster 3 (green) = central feminine–neutral group.

In Taiwan (Figure 3, Figure 5), the cluster composition differs meaningfully. The masculine-adjacent cluster expands to five members (P10 Aesop, P7, P12, P2, P8 MUJI)—notably absorbing MUJI and Aesop, the two earth-tone exemplars that Thai respondents perceived as maximally neutral. P1 Prada Candy forms a singleton cluster, isolated by its saturated pink colouring. The remaining central cluster (orange: P3, P4, P5, P6, P9, P11) gathers the grayscale and pale-toned feminine–neutral designs that constitute Taiwan's neutral core — dominated by the clinical-minimalist exemplars (The Ordinary, CK One, Viktor & Rolf). This clustering divergence directly

visualises the GNS ranking disagreement reported in Section 4.3: earth-tone warmth reads as neutral-organic in Thailand but shifts toward masculine-coded in Taiwan.

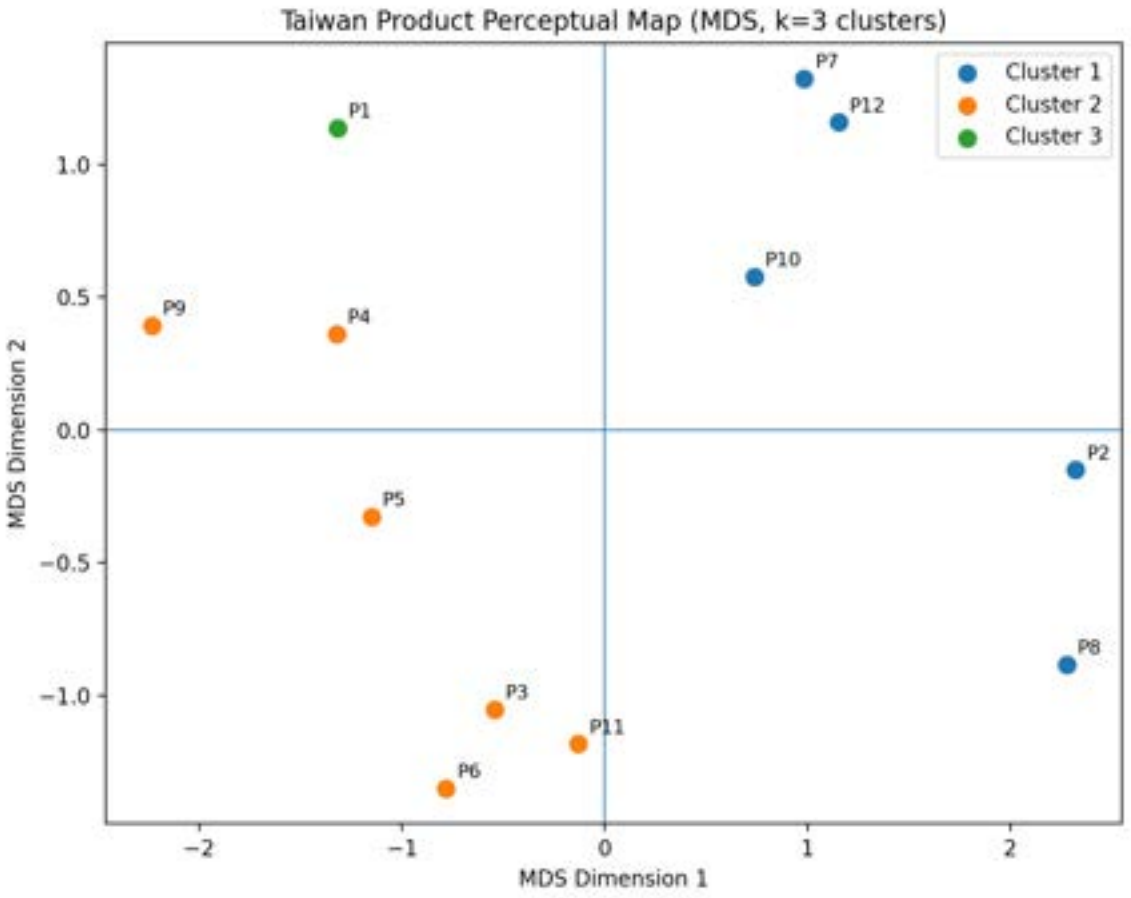


Figure 3. Taiwan product perceptual map (MDS, k = 3 clusters). The masculine-adjacent cluster (blue) absorbs MUJI (P8) and Aesop (P10); Prada Candy (P1) is a perceptual singleton.

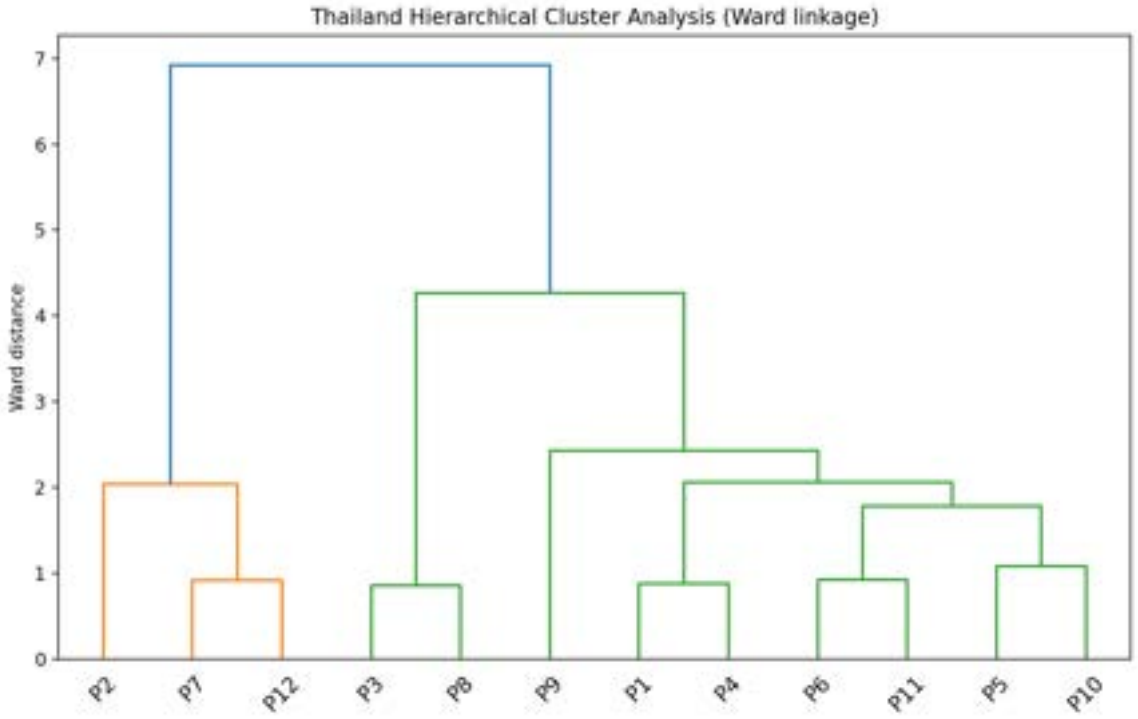


Figure 4. Thailand hierarchical cluster analysis (Ward linkage). The masculine trio (P2, P7, P12) separates at maximal distance; P3 and P8 form a tight minimalist-neutral pair.

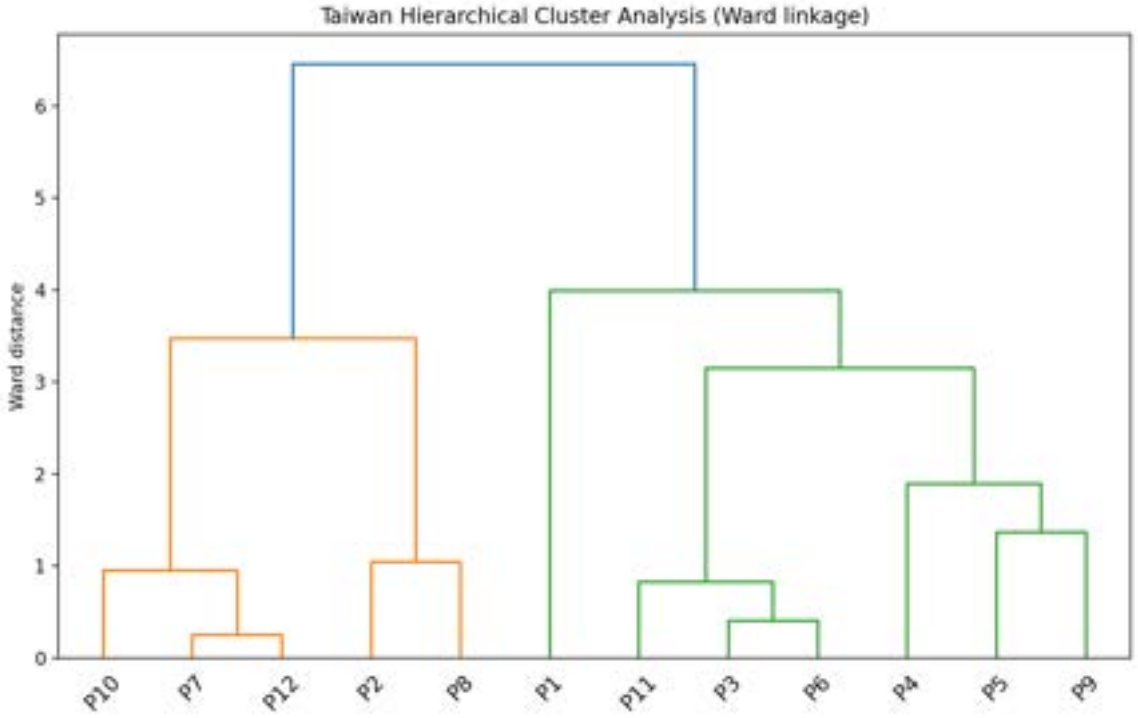


Figure 5. Taiwan hierarchical cluster analysis (Ward linkage). The left branch (P10, P7, P12, P2, P8) groups masculine and earth-tone designs together, diverging from the Thai solution.

The cross-market comparison of image spaces yields a central design insight: the neutral zone is not culturally fixed. Designs occupying the neutral core in one market (MUJI, Aesop in Thailand) can drift toward gendered territory in another (Taiwan), driven primarily by differential colour associations. Grayscale-based designs (P3,

P5, P11) remain within the neutral region in both maps, making grayscale the most robust cross-cultural neutral strategy—consistent with the QT Type I results in Section 4.4.

4.3 Gender-Neutrality Scores and Rankings

Figure 6 presents GNS for all twelve stimuli. The instrument demonstrated strong discriminant validity: expert-validated neutral exemplars (MUJI, Aesop, The Ordinary, CK One) achieved the highest scores, while masculine-coded designs (Paco Rabanne Invictus, Dior Sauvage) anchored the low end. P12 Paco Rabanne Invictus received the lowest GNS in Thailand (6.445), confirming its masculine classification. The Spearman rank-order correlation between Thai and Taiwanese GNS rankings was rs = .32 (N = 12), indicating moderate cross-cultural agreement with meaningful divergence—particularly for MUJI (TH rank 1, TW rank 8) and Prada Candy (TH rank 5, TW rank 12).

[Figure 6. Gender-Neutrality Scores for all 12 packaging stimuli: Thailand (blue) and Taiwan (orange). Dashed red line = GNS 7.0.]

Table 4. Gender-Neutrality Scores by packaging stimulus (N = 12, complete dataset)

ID	Design	TH M(F–M)	TH GNS	TW M(F–M)	TW GNS	TH/TW Rank	Cat.
P8	MUJI Container	4.026	7.929	4.990	6.912	1 / 8	Neutral
P10	Aesop Departure	4.013	7.872	4.451	7.250	2 / 4	Neutral
P9	Viktor & Rolf	3.705	7.840	3.000	7.108	3 / 7	Neutral
P5	The Ordinary Set Box	3.718	7.833	3.255	7.387	4 / 3	Neutral
P1	Prada Candy	3.833	7.705	2.520	6.593	5 / 12	Feminine
P11	CK One	3.705	7.583	3.098	7.549	6 / 1	Neutral
P3	The Ordinary Niacinamide	3.769	7.538	3.245	7.456	7 / 2	Neutral
P4	Lancôme	3.423	7.532	2.667	7.127	8 / 6	Feminine
P6	Dior J'adore	3.372	7.109	2.873	7.186	9 / 5	Feminine
P7	Dior Sauvage	4.936	6.795	5.108	6.765	10 / 9	Masculine
P2	Diptyque Eau Capitale	5.000	6.744	5.284	6.637	11 / 11	Neutral
P12	Paco Rabanne Invictus	5.178	6.445	5.238	6.728	12 / 10	Masculine
	Sample mean GNS	—	7.410	—	7.058	—	—

Note. GNS = 8 − (|M(Delicate–Strong) − 4| + |M(Feminine–Masculine) − 4|) ÷ 2. Max = 8.00 (perfect neutrality). Cat. = expert-validated category.

4.4 Quantification Theory Type I: Importance Weights and Category Scores

QT Type I analysis identified the relative contribution of each design element to GNS. Figure 7 presents importance weight distributions for both samples. Color was the dominant element in both cultures (TH: 42.8%; TW: 32.6%). Font Type ranked second in Thailand (28.2%) but fourth in Taiwan (20.4%), while Surface Finish ranked second in Taiwan (25.5%) and Font Weight was notably more salient in Taiwan (21.5%) than Thailand (12.9%). Table 5 summarises the cross-cultural comparison.

[Figure 7. QT Type I importance weight distributions: Thailand (left) and Taiwan (right), N = 12.]

Table 5. Cross-cultural comparison of QT Type I importance weights (N = 12)

Design element	TH rank	TH range	TH imp.	TW rank	TW range	TW imp.	Pattern

Color (primary)	#1	1.335	42.8%	#1	0.782	32.6%	Shared #1
Font type	#2	0.881	28.2%	#4	0.490	20.4%	TH > TW
Surface finish	#3	0.504	16.1%	#2	0.611	25.5%	TW > TH
Font weight	#4	0.401	12.9%	#3	0.517	21.5%	TW > TH

Note. Importance weight = element range ÷ sum of all ranges × 100%. TH = Thailand (n = 78); TW = Taiwan (n = 102).

Figure 8 shows category-level mean GNS scores. For color, Earth/Kraft achieved the highest neutrality in Thailand (M = 7.929, via MUJI) while Grayscale performed best in Taiwan (M ≈ 7.45, via CK One and The Ordinary). Blue and Black received the lowest GNS in both samples. For surface finish, Matte outperformed Glossy in both cultures, with the gap larger in Taiwan (range = 0.611) than Thailand (range = 0.504). For typography, Sans-serif consistently outperformed Script. Critically, Bold weight achieved higher GNS than Regular in Taiwan (Bold: 7.320 vs. Regular: 6.803), a counterintuitive finding consistent with the Taiwan cluster structure, in which assertive, high-contrast designs are assimilated into neutral territory rather than gendered territory.

[Figure 8. Category-level mean GNS by design element — color (left), typography and surface finish (right), N = 12.]

Table 6. Category-level mean GNS scores by design element and country

Element	Category	TH mean GNS	TW mean GNS	Direction
Color	Earth / Kraft	7.88	6.91	TH >> TW
	Grayscale	7.72	7.45	Both high
	Cream / White	7.49	7.22	Both high
	Pale Pink	7.53	7.13	TH > TW
	Blue	6.80	6.77	Both low
	Black	6.59	6.68	Both low
Surface	Matte	7.71	7.33	Both high
	Semi-gloss	7.34	7.01	Mid
	Glossy	7.21	6.72	Both lower
Font type	Sans-serif	7.57	7.19	Both high
	Serif	7.32	7.16	Mid
	Script	6.78	6.96	Both low
Font weight	Bold	7.59	7.32	Both high
	Light	7.58	7.14	Both high
	Regular	7.36	6.80	TW penalises Regular

Note. Values are mean GNS averaged across all stimuli sharing that category. Ranges may differ from Table 5 due to averaging across unequal cell counts.

4.5 Design Rule: What Raises vs. Lowers Neutrality

Combining color and surface finish coefficients, a clear heuristic emerges. Table 7 presents the QT Type I category scores (β) for key design elements across both markets.

Table 7. QT Type I category scores (β) for key design elements — raises vs. lowers GNS

Direction	Element	TW β	TH β
↑ Raises neutrality	White / Cream	+ 0.32	+ 0.29
	Earth / Kraft	+ 0.02	+ 0.49
	Matte finish	+ 0.22	+ 0.33
	Textured finish	+ 0.19	+ 0.46
↓ Lowers neutrality	Blue	−0.29	−0.61
	Black	−0.23	−0.50
	Pink	−0.17	−0.00
	Glossy finish	−0.22	−0.35

Note. β = category score from QT Type I regression. Positive = increases GNS; negative = decreases GNS. Grand mean GNS (TH) = 7.410; (TW) = 7.058.

The design heuristic derived from these coefficients: White / Earth-Kraft + Matte / Textured + Sans-serif = highest achievable neutrality across both markets. This combination is empirically confirmed by the top-ranking stimuli: MUJI (Earth + Matte + Sans-serif, TH GNS = 7.929), CK One (Grayscale + Matte + Sans-serif Light, TW GNS = 7.549), and The Ordinary (Grayscale + Matte + Sans-serif Bold, TW GNS = 7.456).

4.6 Answering the Research Questions

Table 8. Research question responses derived from QT Type I findings

Objective	RQ	Answer
IDENTIFY	RQ1.1 Which elements?	Color and Surface Finish correlate most strongly with neutrality in both cultures.
	RQ1.2 Relative importance?	Color ≈ 0.6–0.7 › Finish ≈ 0.5 › Font Weight ≈ 0.2 › Font Type < 0.1 (partial η^2)
MEASURE	RQ2.1 GNS range?	GNS ranges ≈ 6.4–7.9; expert-validated neutral exemplars consistently score highest.
	RQ2.2 Optimal combination?	White / Earth-Kraft + Matte / Textured + Sans-serif maximises GNS in both markets.
APPLY	RQ3.1 Design principles?	Restrained chroma, non-reflective surface, minimal ornament — confirmed cross-culturally.
	RQ3.2 Predict before build?	Yes — the QT-I equation predicts GNS from design specifications before a prototype is built.

4.7 Generational Subgroup Analysis (Gen Z vs. Gen Y)

To examine whether the findings generalise across consumer generations, all analyses were repeated separately for Generation Z (aged 30 or below) and Generation Y (31 and above) within each country. The Taiwanese subgroups (Gen Z n = 56; Gen Y n = 46) partition the main sample exactly; the Thai generational analysis uses all complete Thai responses (N = 115; Gen Z n = 67, Gen Y n = 48), a superset of the main sample, and is therefore reported as a robustness analysis. Three findings emerged.

First, generational agreement diverges sharply by country. In Taiwan, Gen Z and Gen Y GNS rankings were nearly identical (Spearman rs = .95, $p < .001$), and no stimulus showed a significant generational difference on the Feminine–Masculine dimension (all $p > .14$). In Thailand, agreement was only moderate (rs = .60, p = .041): Thai Gen Y placed Prada Candy exactly at the scale midpoint (M F–M = 4.00; GNS rank 1), whereas Gen Z perceived it as further from neutrality (M = 4.60; rank 9) — an eight-rank shift — and Gen Z penalised masculine-coded designs more severely (Diptyque F–M: Gen Z 5.27 vs. Gen Y 4.69, p = .007).

Second, element salience shows a cross-national generational pattern (Table 9, Figure 9). Color ranks first in every country × generation cell, confirming its universal dominance. Surface finish, however, is comparably salient for Gen Z in both countries (Thailand 24.0%; Taiwan 25.9%) while dropping for Thai Gen Y (14.7%). This suggests that part of the national difference in surface salience reported in Section 4.4 reflects generational composition rather than culture alone: surface sensitivity appears to be a Generation Z characteristic that transcends the two markets.

Third, cultural effects persist within every generation. Earth tones carry negative category scores in Taiwan for both generations (Gen Z $\beta = -0.10$; Gen Y $\beta = -0.14$), confirming that the MUJI ranking divergence between countries is cultural rather than generational. Conversely, the clinical-minimalist profile is robust across all segments: The Ordinary Set Box (grayscale + sans-serif + matte) is the only stimulus ranked in the top five of all four country × generation cells, and matte finish is the only design category with a positive coefficient in every cell — making grayscale-plus-matte the most segment-proof gender-neutral configuration in the dataset.

Table 9. QT Type I importance weights by country and generation (2 × 2 matrix)

Design element	TH Gen Z	TH Gen Y	TW Gen Z	TW Gen Y	Pattern
Color (primary)	50.4%	49.2%	33.4%	55.8%	#1 in every cell
Surface finish	24.0%	14.7%	25.9%	22.0%	Gen Z > Gen Y in both countries
Font type	17.5%	25.6%	18.6%	12.8%	Salient only for TH Gen Y
Font weight	8.2%	10.5%	22.2%	9.4%	Salient only for TW Gen Z

Note. TH generational analysis based on all complete Thai responses (N = 115); TW subgroups partition the main sample (N = 102). Within-generation Spearman agreement: Taiwan rs = .95; Thailand rs = .60.

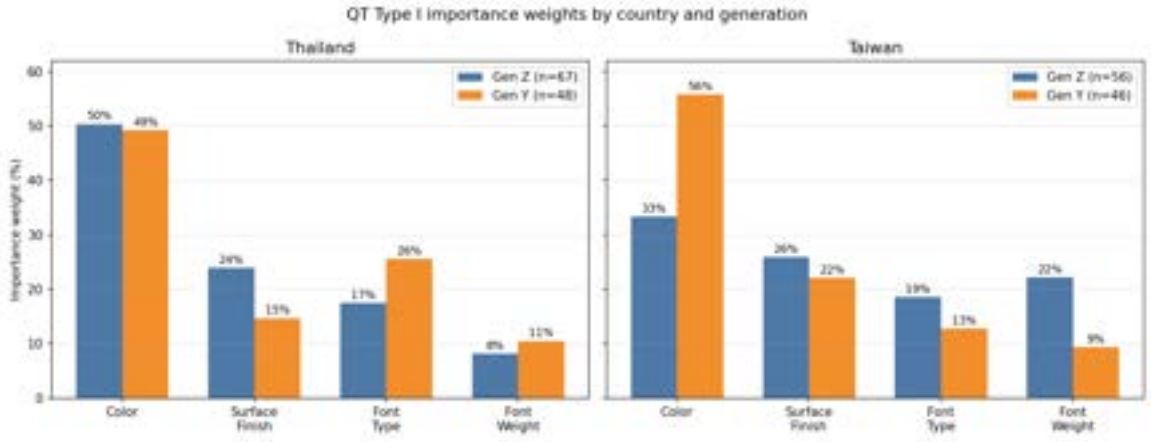


Figure 9. QT Type I importance weights by country and generation. Color dominates every cell; surface-finish salience characterises Gen Z in both markets.

5. DISCUSSION

5.1 Cross-Cultural Perceptual Architecture

The MDS perceptual maps and Ward cluster solutions reveal that the overall organisation of the perceptual space is similar across cultures—gendered-masculine designs separate first and at the greatest distance in both markets—but the membership of the neutral core diverges. In Thailand, earth-tone designs (MUJI, Aesop) sit at the centre of the neutral territory, whereas in Taiwan the same designs migrate toward the masculine-adjacent cluster, and the neutral core is held by grayscale, clinical-minimalist designs (CK One, The Ordinary). This confirms Orth and Malkewitz's (2008) holistic packaging theory whilst demonstrating that the holistic impression itself is culturally

calibrated: identical design configurations are assimilated into different perceptual neighbourhoods depending on culturally learned colour associations.

5.2 Color, Typography, and Surface Findings

Primary color dominates gender-neutral perception in both cultures (TH: 42.8%; TW: 32.6%), confirming van Esch et al.'s (2020) finding that color is the most immediate gender-coding cue. Earth tones rank highest in Thailand (MUJI, M = 7.929) whilst Grayscale ranks highest in Taiwan (CK One, M = 7.549), suggesting differential cultural associations with organic warmth versus clinical minimalism. Pink and Black received the lowest GNS in both samples—Pink consistently gendered-feminine and Black gendered-masculine across cultures.

Font Type is more salient in Thailand (28.2%) than Taiwan (20.4%), potentially related to the Thai sample's substantially higher proportion of gender-diverse respondents (51.3% non-binary vs. 1% in Taiwan), who may apply more stringent evaluation criteria for font-based gender signals. Surface Finish is more influential in Taiwan (25.5%), consistent with Taiwanese consumers integrating material tactility into their gender perception schema. The counterintuitive finding that Bold typography achieves higher GNS than Regular in Taiwan suggests that, for Taiwanese consumers, typographic assertiveness is partially decoupled from gender coding—assertive but chromatically restrained designs remain within neutral perceptual territory.

5.3 Practical Design Guidelines (Priority-Ordered)

G1 — Color (Essential, ≥ 42% importance in both cultures): Use grayscale, cream/white, or off-white across both markets. For Thai markets, earth tones (kraft, warm beige) achieve the highest GNS (≈ 7.93). Avoid pink and black as primary colors—both received the lowest GNS in both samples.

G2 — Surface Finish (Critical, especially Taiwan 25.5%): Specify matte as default. The matte–glossy gap is larger in Taiwan (range 0.611) than Thailand (range 0.504). Textured / uncoated kraft surfaces perform strongly in Thailand.

G3 — Font Type (Critical, especially Thailand 28.2%): Default to geometric or grotesque sans-serif. Script fonts consistently received the lowest GNS in both cultures and should be avoided for gender-neutral positioning.

G4 — Font Weight (Important, especially Taiwan 21.5%): Bold and Light weights achieve higher GNS than Regular in Taiwan (Bold: 7.320; Light: 7.135 vs. Regular: 6.803). Regular weight should be used cautiously in Taiwanese markets.

G5 — Localisation: Taiwan is colour-sensitive with greater surface finish awareness; Thailand penalises Black and Glossy more severely. Country moderates the relative salience of each element.

Figures 10a–10c illustrate the guidelines applied end-to-end: three market-calibrated cartons for the TINSPIRE case study introduced in Section 6.4, each operationalising G1–G5 for its target market.

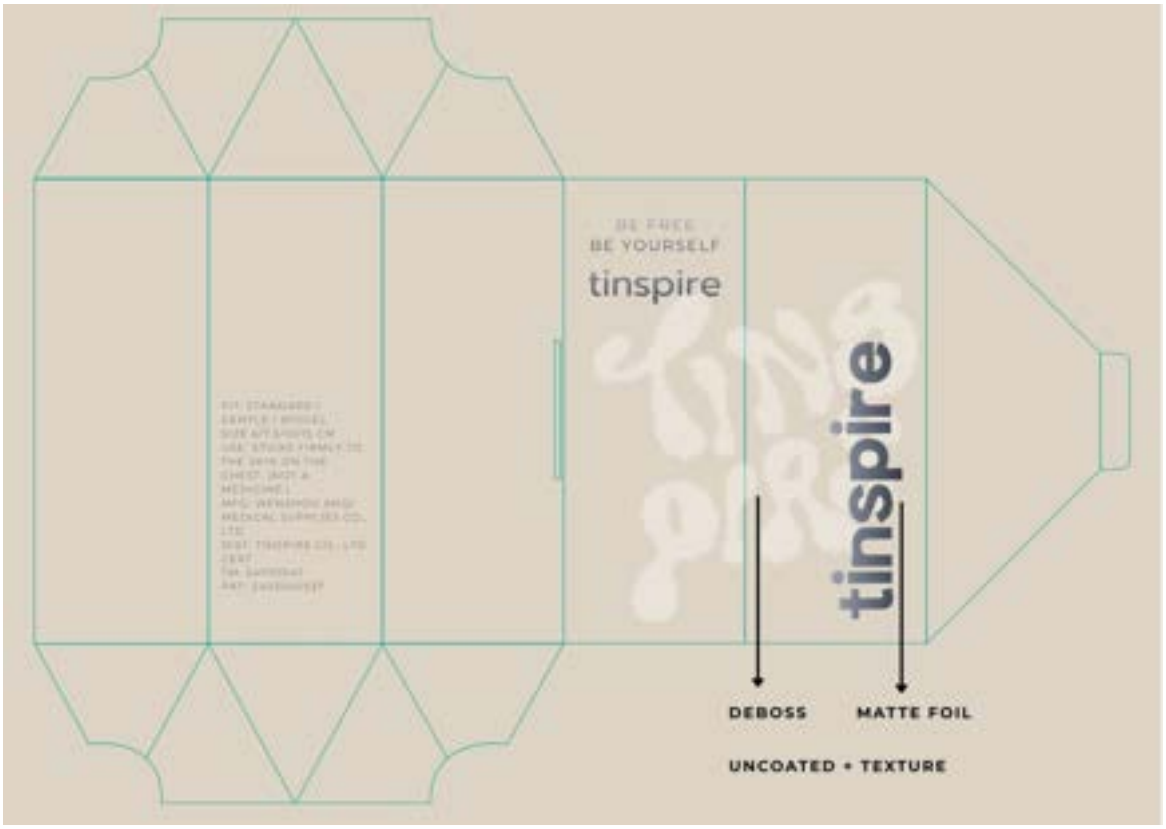


Figure 10a. TINSPIRE Thailand version: kraft base, uncoated textured stock, debossed flat wordmark, matte foil.

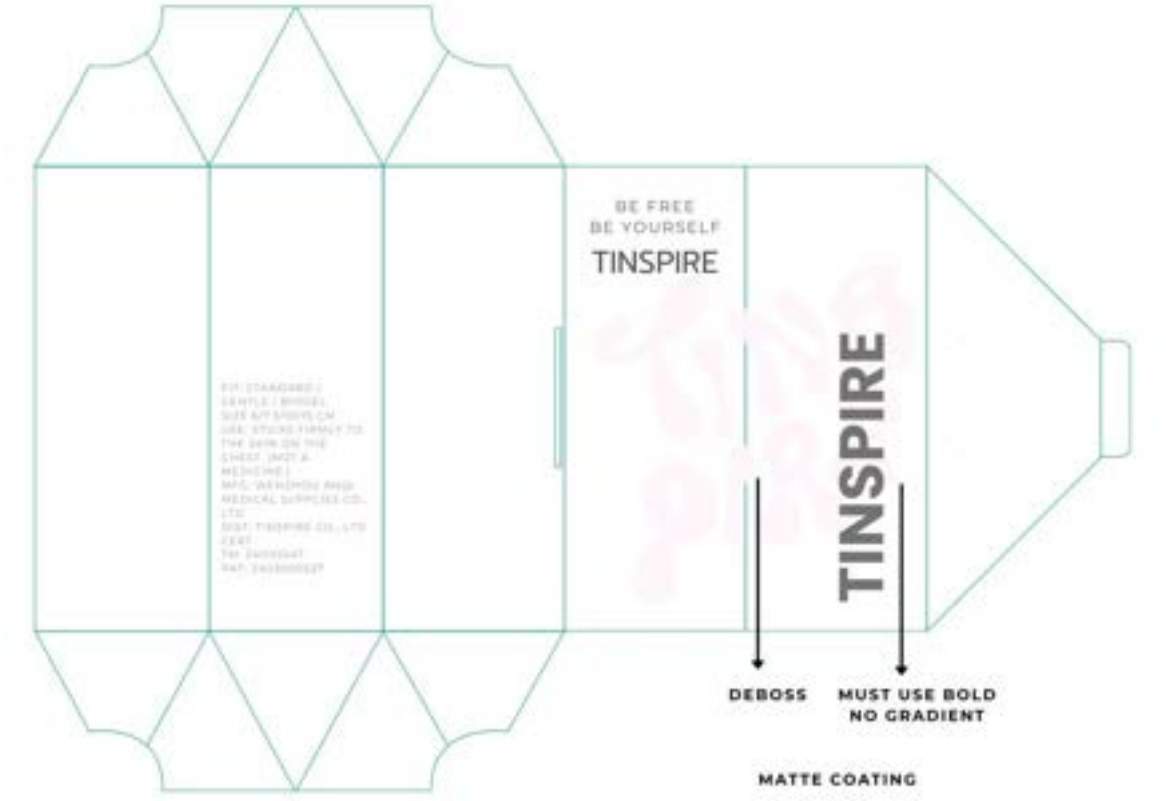


Figure 10b. TINSPIRE Taiwan version: white clinical base, Bold flat wordmark (no gradient), matte coating, deboss.

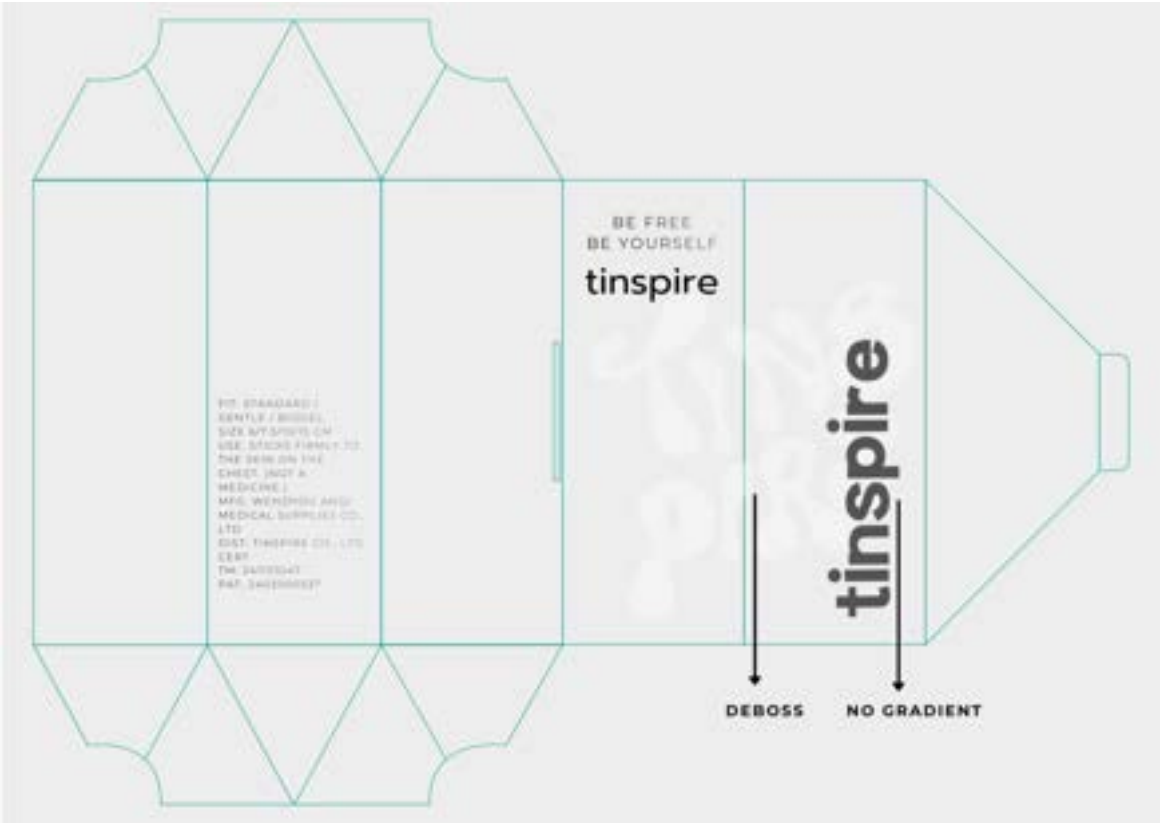


Figure 10c. TINSPIRE Universal version: grayscale base, flat charcoal wordmark, deboss, no gradient.

5.4 Generational Moderation of Element Salience

The subgroup analysis qualifies the cross-cultural findings in two ways. Practically, a single gender-neutral design can serve all age segments in Taiwan (within-country generational agreement rs = .95), whereas the Thai market may warrant generational calibration: Thai Gen Z applies stricter criteria to masculine-coded cues and still reads saturated pink as feminine, while Thai Gen Y evaluates pink at the neutral midpoint. Theoretically, the analysis decomposes the national surface-finish difference: because Gen Z surface salience is nearly identical across countries, the lower aggregate Thai figure is driven by Thai Gen Y — indicating that some element-salience differences attributed to culture are partly compositional. One plausible mechanism for the wider Thai generational gap is the concentration of gender-diverse respondents (51.3% non-binary) in the Thai sample, predominantly Gen Z, who may apply more stringent gender-signal criteria; formally modelling generation and gender identity as moderators is a priority for future work.

6. DESIGN PROTOTYPES AND VALIDATION

6.1 Validation Against Real Exemplars

Before proposing new prototypes, the guidelines were validated against the four highest-ranking real exemplars in the dataset. Table 10 confirms that all four satisfy G1 (neutral/earth color) and G2 (matte/textured surface) and achieve high GNS in at least one market, providing empirical grounding for the guidelines.

Table 10. Validation of design guidelines against top-ranked real exemplars

Brand	Color / Finish	GNS TW	GNS TH	Guidelines satisfied
CK One	White / Matte	7.55	7.58	G1 ✓ G2 ✓ G3 ✓ G4 ✓

The Ordinary	Grayscale / Matte	7.46	7.54	G1 ✓ G2 ✓ G3 ✓ G4 ✓
Aesop	Earth / Textured	7.25	7.87	G1 ✓ G2 ✓ G3 ✓
MUJI	Earth / Matte	6.91	7.93	G1 ✓ G2 ✓ G3 ✓

Note. All four exemplars satisfy G1 and G2 and score GNS ≥ 7.0 in at least one market, confirming guidelines describe what already works in practice.

6.2 Forward Prediction: New Design Configuration

The QT Type I equation enables prediction of GNS for untested configurations before prototyping. Example: a new skincare carton specified as White + Matte + Sans-serif + Light weight. Predicted GNS = Grand mean (7.06) + White (+0.32) + Matte (+0.22) + Sans-serif (+0.07) + Light (+0.31) = 7.98. This exceeds all existing stimuli and represents the theoretical ceiling for a cross-market neutral design under current guidelines.

To make the prediction model directly usable by practitioners without statistical training, it was implemented as an interactive tool, the GNS Analyzer (Figure 11). The practitioner uploads a packaging image (front, side, or flat-lay view); the design elements identified in the image—primary colour, typography, and surface finish—are mapped to the corresponding QT Type I category scores, and the tool returns a predicted Gender-Neutrality Score positioned on a five-band reference scale anchored by the validated stimuli (Highly Neutral ≥ 7.80, exemplified by MUJI and Aesop, through Gendered < 7.00, exemplified by Paco Rabanne and Dior Sauvage). This converts the regression model into a design-stage screening aid: candidate packaging can be evaluated and compared before any prototype is produced. The reference scale also communicates a base-rate insight from the dataset—approximately 80% of commercial packaging scores below GNS 7.0, underscoring that measured neutrality remains rare in the market.

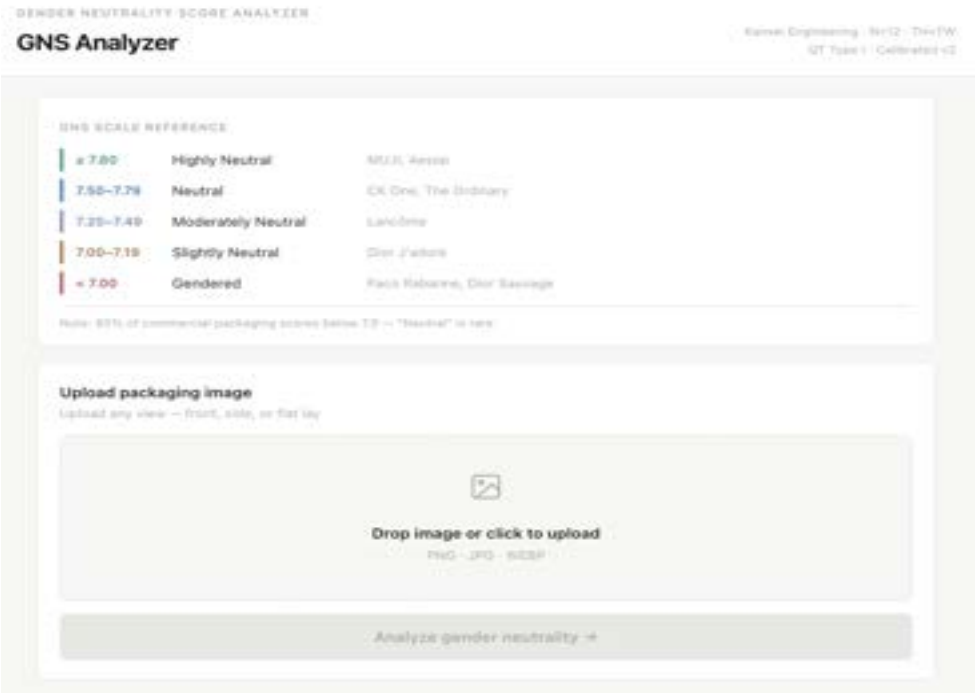


Figure 11. The GNS Analyzer: an interactive implementation of the QT Type I model. A packaging image is uploaded and scored against a five-band reference scale anchored by the validated exemplars.

6.3 Three Design Prototypes

Three packaging prototypes operationalise the QT Type I guidelines as concrete design concepts, each targeting a specific cultural market configuration. Figure 12 illustrates all three prototypes.

[Figure 12. Design Prototypes A (Universal Minimal), B (Earth Warmth, TH-optimised), and C (Clinical, TW-optimised).]

Table 11. Design prototype specification matrix and expected GNS

Specification	Prototype A: Universal Minimal	Prototype B: Earth Warmth (TH)	Prototype C: Clinical (TW)
Primary color	Mid-grey (#808080)	Kraft brown (#E8DDD0)	Pure white (#FFFFFF)
Color category	Grayscale	Earth tone	Cream/White
Typography	Geometric sans-serif	Sans-serif	Bold geometric sans-serif
Font weight	Bold + Regular mix	Regular	Bold primary / Light secondary
Surface finish	Matte	Matte (uncoated kraft)	Matte
Visual complexity	None — pure typography	Minimal (line rules only)	None — typography only
Expected GNS (TH)	~7.80	~7.90	~7.60
Expected GNS (TW)	~7.40	~7.00	~7.50
Primary evidence	P5+P11 cross-cultural consensus	P8 MUJI — TH rank 1	P11+P3 — TW ranks 1 and 2
Market target	TH + TW cross-market	Thailand-primary	Taiwan-primary

6.4 Industry Case Study: TINSPIRE Market-Calibrated Packaging

To demonstrate end-to-end applicability, the guidelines and generational findings were applied to a real product development case: TINSPIRE, a gender-neutral chest-tape (underwear accessory) brand preparing carton packaging for the Thai and Taiwanese markets. Three market-calibrated cartons were designed on a shared structural dieline, varying only the design elements identified by the QT Type I models (Figures 10a–10c, Section 5.3).

Thailand version (kraft). An earth/kraft base was selected because Earth is Thailand's strongest colour category and resonates most with Thai Gen Z (subgroup $\beta = +0.12$ vs. Gen Y +0.02). Print specification exploits Thailand's texture premium ($\beta = +0.46$): uncoated textured stock, debossed wordmark, and matte foil in place of gloss foil (avoiding the Thai glossy penalty, $\beta = -0.35$). The logotype is a single flat colour — the gradient used in an earlier draft was removed, since metallic-gradient treatments belong to the gendered luxury vocabulary (cf. Blue-Gold, the most negative colour category in the pilot Thai study).

Taiwan version (white / clinical). A white base with no background pattern follows the clinical-minimalist archetype that dominates Taiwanese rankings. The wordmark is set in Bold with no gradient, reflecting two Taiwan-specific findings: Regular weight is penalised in both Taiwanese generations ($\beta = -0.13$ / -0.07), and Taiwanese consumers integrate dynamism into gender evaluation, making gradients riskier in this market than in Thailand. Matte coating is specified as the default finish (Taiwan surface salience 25.5%).

Universal version (grayscale). A light-grey base carries a flat charcoal wordmark. Grayscale is the only colour category with positive coefficients in all four country × generation cells (+0.23 / +0.10 / +0.22 / +0.35), and the profile replicates The Ordinary Set Box — the only stimulus ranked top-five in every cell. A dual-weight lockup (Bold primary, Light secondary) averages the opposing weight preferences of the two markets (Regular best in Thailand, worst in Taiwan).

Applying each subgroup's QT Type I model to the three specifications yields the twelve forward predictions in Table 12. The Thai market is permissive — all three versions reach or approach the scale ceiling in both Thai generations — whereas Taiwan discriminates: the kraft version underperforms for Taiwanese Gen Y (predicted

GNS = 6.90), confirming it as a Thailand-exclusive SKU, while the grayscale Universal version is the strongest Taiwanese performer in both generations and the only single-SKU option supported across all four cells.

Table 12. Forward-predicted GNS of the three case-study cartons by country × generation subgroup model

Version	TH Gen Z	TH Gen Y	TW Gen Z	TW Gen Y	Deployment
Kraft (TH)	ceiling (8.0)	ceiling (8.0)	7.09	6.90	Thailand-exclusive
White (TW)	ceiling (8.0)	ceiling (8.0)	7.45	7.21	Taiwan-primary
Grayscale (Universal)	ceiling (8.0)	7.93	7.54	7.47	Single cross-market SKU

Note. Predictions from subgroup QT Type I models (N = 12 stimuli per model); Thai raw predictions exceeding the 8.00 scale maximum are reported as ceiling. Untested elements (muted accent hues) are excluded from the prediction and flagged for follow-up validation with the same four-scale instrument.

7. CONCLUSIONS

This study provides the first cross-cultural empirical analysis of design element importance weights for gender-neutral packaging perception, with a complete N = 12 stimulus set including both neutral and strongly-gendered reference designs. Four contributions are reported.

First, primary color is confirmed as the universally dominant element (TH: 42.8%; TW: 32.6%), with grayscale and earth tones as most reliable neutral choices and pink/black as reliably gendered. Second, cross-cultural divergences in element salience—Font Type critical in Thailand (28.2%), Surface Finish and Font Weight critical in Taiwan (25.5% and 21.5%)—require culturally calibrated design decisions. Third, the complete N = 12 dataset, including P12 Paco Rabanne Invictus as a Masculine reference design, strengthens the discriminant validity of the GNS instrument by anchoring the low end of the neutrality scale with an expert-validated masculine stimulus. Fourth, the 'clinical-minimalist' archetype—neutral palette, sans-serif typography, matte surface—consistently achieves the highest cross-cultural gender-neutrality, whilst the clinical prototype achieves the strongest Taiwan performance. Fifth, a generational subgroup analysis shows that this archetype is also segment-proof: rankings converge across generations in Taiwan (rs = .95) and, although Thai generations diverge (rs = .60), the grayscale–sans-serif–matte configuration remains top-ranked in every country × generation cell, while surface-finish salience emerges as a Generation Z characteristic common to both markets.

Future research should extend the cross-cultural comparison to additional Asian and Western markets, formally test country × generation interactions with mixed-effects models, explicitly model gender identity as a moderating variable in QT Type I analysis, and validate findings with physical packaging prototypes incorporating tactile and three-dimensional properties.

REFERENCES

Grand View Research. (2023). Gender-neutral cosmetics market size, share & trends analysis report 2023–2030. Grand View Research.

Grohmann, B. (2009). Gender dimensions of brand personality. Journal of Marketing Research, 46(1), 105–119. https://doi.org/10.1509/jmkr.46.1.105

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. Annals of the Institute of Statistical Mathematics, 3(2), 69–98.

Henderson, P. W., Giese, J. L., & Cote, J. A. (2004). Impression management using typeface design. Journal of Marketing, 68(4), 60–72.

Lieven, T., Grohmann, B., Herrmann, A., Landwehr, J. R., & van Tilburg, M. (2014). The effect of brand gender on brand equity. Psychology & Marketing, 31(5), 371–385.

Lieven, T., Grohmann, B., Herrmann, A., Landwehr, J. R., & van Tilburg, M. (2015). The effect of brand gender on brand equity. Psychology & Marketing, 32(6), 638–651.

Mintel. (2022). Beauty and personal care market trends: Gender neutrality. Mintel Group Ltd.

Murata, A. (2019). Kansei engineering—A perspective from product and system design. Applied Ergonomics, 74, 112–126.

Nagamachi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. Applied Ergonomics, 33(3), 289–294.

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books.

Orth, U. R., & Malkewitz, K. (2008). Holistic package design and consumer brand impressions. Journal of Marketing, 72(3), 64–81.

van Esch, P., Arli, D., Ghasemaghaei, M., Blut, M., & Mende, M. (2020). Designing for the genders: The role of visual harmony. International Journal of Research in Marketing, 37(4), 766–785.

Wolin, L. D. (2003). Gender issues in advertising: An oversight synthesis of research. Journal of Advertising Research, 43(1), 111–129.

Cross-Cultural Interpretation of Kansei Vocabulary for Virtual Nature

Petra NURMELA[1a], Emma KIRJAVAINEN[a], Jonna HÄKKILÄ[a]

[a] *University of Lapland, Finland, firstname.lastname@ulapland.fi*

Abstract

This study examines how Finnish, Japanese, and English speakers interpret Kansei words in the context of Virtual Nature Experiences (VNE). Through three participatory workshops (n = 20) and a survey, we explored differences between participants' ratings of how well Kansei words described a VNE and their broader emotional associations with those words, particularly for terms such as *Peaceful* and *Nostalgic.* Participants' associations emphasised nature, tranquillity, and experiential qualities, while suggesting that multilingual Kansei words were not always understood in the same way by participants. These findings highlight challenges in selecting and validating multilingual Kansei vocabularies for evaluation tools. The study provides an exploratory foundation for developing a culturally responsive Kansei Evaluation Tool to capture emotional and sensory perceptions of VNE more effectively.

Keywords: *Vocabulary Evaluation, Cross-Cultural Perception, Virtual Nature, Kansei Design*

1 INTRODUCTION

As urbanisation expands, reduced access to natural environments has become a global concern, impacting well-being and psychological health (Larson et al., 2019). Everyday interactions with nature are declining, particularly among younger generations, contributing to stress and diminished mental resilience (Kaplan & Kaplan, 1989; Ulrich, 1981). Especially, how individuals perceive their relationship with the natural world, i.e. Nature Connectedness (NC), is strongly linked to well-being and life satisfaction (Capaldi et al., 2014). While physical nature contact remains essential, virtual technologies are increasingly explored as tools for fostering these connections. Immersive environments, such as 360° videos and computer-generated landscapes, can promote NC when they provide sensory richness and interactivity (Calogiuri et al., 2023). For instance, enhancements such as haptic feedback further improve vitality and mood (Nurmela et al., 2025a). However, existing evaluation methods for VNE are often generic, lengthy, and culturally insensitive, limiting their applicability across diverse user groups.

---

[1] Yliopistonkatu 8, 96300 Rovaniemi, Finland

This study addresses a critical gap by developing a structured, culturally responsive tool for evaluating VNE. Its primary aim is to explore how participants interpret and emotionally respond to ten previously identified low-level Kansei words (Nurmela et al., 2025b) within multilingual contexts of VN. By applying Kansei-based methodologies, this work advances the KEER discourse and demonstrates its potential to enhance cross-cultural design evaluation in emerging domains such as VN.

2 KANSEI DESIGN-BASED EVALUATION TOOL FOR VIRTUAL NATURE

Kansei is a Japanese concept that encompasses emotions, sensitivity, aesthetics, intuition, and sensory experience (Harada, 1998). Kansei Design (KD) approaches user experience from an affective perspective, emphasising how individuals feel, sense, and interpret stimuli (Nagamachi, 2011). Lee et al. (2002) describe it as reflecting “soft values” within product and spatial design. According to Lokman (2010), Kansei emerges through sensory input (vision, hearing, touch, smell, and taste) processed via emotional and cognitive interpretation. Lévy (2013) further frames KD as a structured methodology for emotional design, integrating these impressions into systematic design processes.

The aim of developing a KD-based evaluation tool is to translate subjective emotional responses to VN into structured and measurable parameters. The process, inspired by the work of Schütte et al. (2024), turns emotional impressions into usable evaluation metrics through four steps: (1) *Domain Definition:* VNE; (2) *Semantic Space Construction:* Kansei words; (3) *Property Space Identification:* visual and interactive attributes of VN; (4) *Testing and Validation* in VN. This paper focuses on step (2) Kansei structure identification in the form of *multilingual workshop*s, where we conduct sessions in Finnish, Japanese, and English to examine how participants interpret and emotionally respond to previously identified Kansei words in the context of VNE (Nurmela et al., 2025b). This stage explores cross-cultural nuances in meaning and affective perception, ensuring that the preliminary vocabulary accurately reflects diverse experiential interpretations before its application in the evaluation tool.

3 STUDY

This section outlines the study and data collection methods, including the participants’ background information, workshops, and the survey.

3.1 Participants

A total of 20 participants took part: nine Finnish speakers, six Japanese speakers, and five English speakers. All were informed of the study procedures at the start of the workshop, and written consent was obtained. Participants were aged 18–24 (n = 8), 25–34 (n = 9), or 35–44 (n = 3). VR familiarity varied: four had never used VR, one had tried it once, twelve had used it 2–5 times, and one had used it more than five times. Spending time in nature was frequent, with 40% visiting weekly, 40% monthly, and 20% daily. Most participants (60%) felt more at home in natural environments; 10% reported a strong preference for nature, 20% preferred a balanced mix of natural and urban environments, and 10% preferred urban settings.

3.2 Workshops

We conducted three participatory workshops, one in Finnish, one in Japanese and one in English. The workshop aimed to evaluate VNE-related preliminary Kansei vocabulary created in previous participatory workshops (Nurmela et al., 2025b). The workshop consisted of three tasks. (1) *Word Association Mapping:* Participants wrote on post-it notes what each word evokes in their mind. (2) *Experience Reflection:* Participants had a short VNE (winter forest walk) and arranged the words on a physical paper scale according to how well the word described their experience. (3) *Multilingual Comparison:* Participants grouped words across languages and discussed the nuances which each word evokes.

3.3 Survey

After the workshop tasks, participants filled out a survey where they were asked to evaluate the same words: (A) *How well the word described their virtual nature experience?* (in the context of the VNE), and (B) *How emotionally connected they felt to this word?* (in general). They evaluated each word using a 5-point Likert scale (1 = not at all; 5 = very well).

4 RESULTS

This section presents the results from the three workshop tasks and the post-session survey with descriptions of the analysis methods.

4.1 Word Association Mapping

The analysis of the data was conducted using the affinity mapping method (Lucero, 2015). The results were analysed by creating an affinity diagram (see Table 1) to categorise the key themes mentioned in the responses. The presented table summarises the thematic categories identified through affinity mapping, a visual technique for organising qualitative data into clusters of semantically related concepts. Affinity mapping enables researchers to group individual data points into meaningful categories, facilitating the identification of patterns and themes across the dataset.

Table 1. The key themes of words evoked by each Kansei word

KANSEI VOCABULARY	THEMES OF WORDS		
Luonto / 自然 / Nature	Nature - Greenness (14)	Silence - Peace (8)	Animals - Environment (6)
Outous / 畏怖 / Strange	Interesting (6)	Scary (5)	Stranger (4)
Rauha / 穏やか / Peaceful	Home - Nature (6)	Safety (4)	Silence (4)
Infrastruktuuri / 街並み / City	Traffic - Noise (7)	Built - Environment (6)	People - Social Spaces (5)
Nostalgia / 郷想 / Nostalgic	Childhood (10)	Memories (8)	Games - Play (6)

Henkisyys / 神秘 / Spirituality	Belief - Beliefs (9)	Experience (5)	Nature (4)
Ympäristö / 環境 / Environment	Nature (7)	Buildings - Environment (5)	Society (3)
Aistillisuus / きれい / Beautiful	Sensitivity (8)	Nature (4)	Aesthetic Experience (4)
Ilmapiiri / 雰囲気 / Atmosphere	Atmosphere - Lived Experience (15)	Air- Space (8)	Senses (4)
Pakoarjesta / 脱日常 / Escapism	Peace - Relaxation (8)	Nature (5)	Travel (5)

The table shows how frequently keywords appeared within broader thematic categories. The most prominent themes relate to Nature and Atmosphere, indicating a strong emphasis on natural elements and experiential qualities in participants' responses. Nostalgic themes such as Childhood and spiritual or belief-related concepts also appear with moderate frequency. In contrast, categories associated with the built environment and urban contexts are present but less prominent. This pattern suggests that participants more often referenced experiences connected to nature, tranquillity, and personal or emotional meaning than those related to urban infrastructure.

4.2 Experience Reflection

The participants were shown a short VNE in the form of a video of a winter forest walk. After the exposure, they arranged words on a physical paper scale based on how well they described their VNE. All the scales were photographed and analysed to capture the average for each language group. Figure 1 presents the distribution of the words across three languages. It can be noticed that most of the words have similar rankings. Although the Japanese group tended to rank these words higher, the small subgroup sizes (n = 5–9) mean these patterns are indicative rather than conclusive. This difference can be seen in the case of *Beautiful,* which is the best-fitting word among the Japanese, whereas it comes after *Nature* and *Environment* in the Finnish and English groups. Similarly, *Atmosphere*, *Spirituality, and Peaceful* are notably higher than their counterparts in other groups. This can be due to the Japanese idea of beauty, where aesthetic experience is related to the holistic experience of the entire surrounding world (Saeki, 2001). These kinds of cultural tendencies, as well as individual differences, are important to take into consideration when designing a tool to capture people's wide perceptions of VN.

4.3 Multilingual Comparison

First, participants matched words across languages individually. After each of them had drawn their own map, they were encouraged to discuss their choices. Many participants pointed out that several word pairs do not carry equivalent meanings. Examples: "Aistillisuus & beautiful doesn't necessarily match in my mind." (P8); "Strange & 畏怖, meaning a bit different." (P15); "nostalgic & 郷想, maybe 懐かしい suit better?" (P17). This outcome suggests

that although the preliminary vocabulary developed during earlier workshop sessions was carefully considered and internally coherent for the original participants, the underlying rationale did not transfer to the new groups. For instance, the set Aistillisuus / きれい / Beautiful does not represent a direct translation sequence but rather an interpretive alignment of meaning, which was not immediately evident to the subsequent participants. Also, some words were viewed as too abstract or too broad, making interpretation difficult, e.g. “Ilmapiiri (atmosphere) can also be an emotional word.” (P13); “With each word, I think about the ambiguity of the word and its meaning to me.” (P10). Participants recognise that several terms carry layered meanings, which may not be stable across users or cultures. These findings indicate the need to refine the vocabulary so that all terms form a coherent conceptual domain and are easily interpreted across languages. Simplifying the terminology by avoiding rare words, complex kanji, and terms with multiple meanings, and favouring adjectives over nouns may help unify meaning across languages.

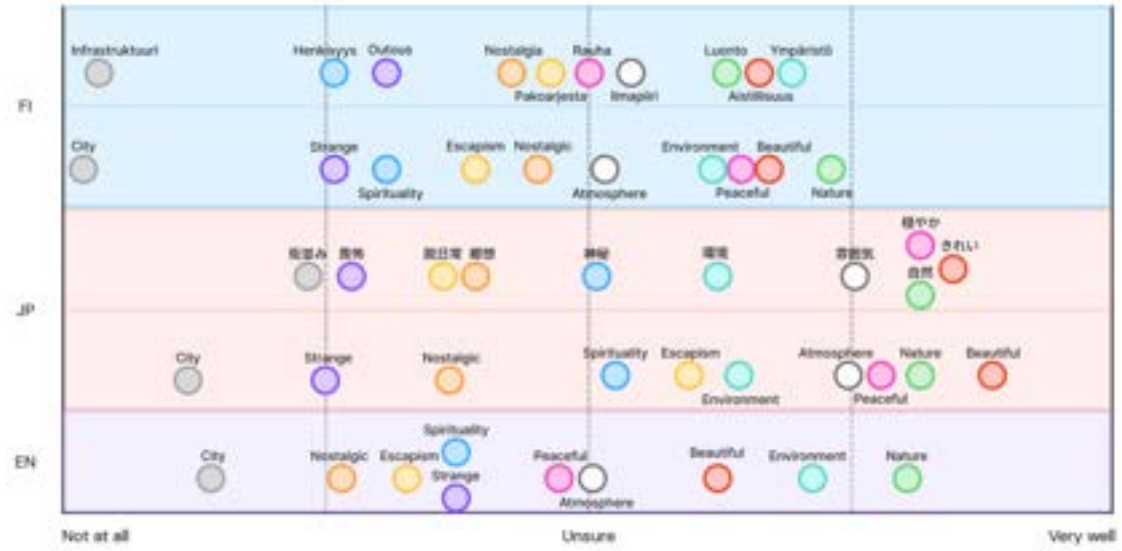


Figure 1. Average ratings of the descriptive accuracy of the words across languages.

4.4 Survey

Figure 2 presents a scatterplot illustrating the ratings of the words in each language by their *Descriptiveness* and *Emotional Connection*. Similarly to the mapping of the words in the Experience Reflection task, it can be seen that a group of Japanese words has higher ratings in descriptiveness: 自然 (Nature), 穏やか (Peaceful), きれい (Beautiful), 雰囲気 (Atmosphere), 環境 (Environment). Whereas Infrastruktuuri and 街並み (City) stay as the least descriptive.

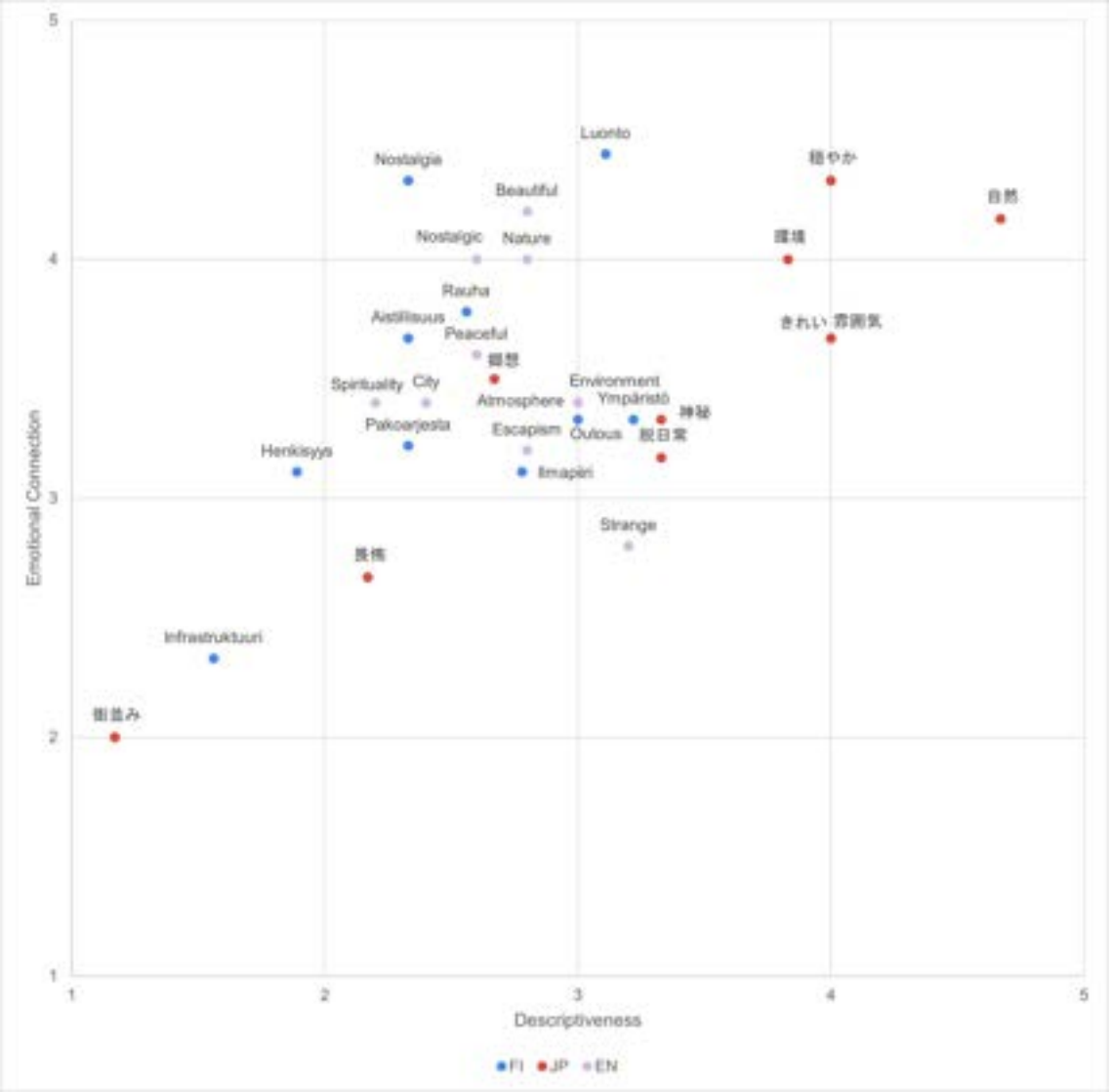


Figure 2. A scatterplot illustrating the ratings of the words in each language by their Descriptiveness and Emotional Connection

Wilcoxon signed-rank tests were conducted for each Kansei word to compare participants' paired Descriptiveness and Emotional Connection ratings. Before correction for multiple comparisons, significant differences were observed for Nature ($Z = -2.280$, $p = .023$), Peaceful ($Z = -3.082$, $p = .002$), City ($Z = -2.675$, $p = .007$), Nostalgic ($Z = -3.335$, $p < .001$), Spirituality ($Z = -2.113$, $p = .035$), and Beautiful ($Z = -2.257$, $p = .024$). No significant differences were found for Strange ($p = .487$), Environment ($p = .563$), Atmosphere ($p = .670$), and Escapism ($p = .246$). After applying the Holm correction for multiple comparisons, only *Peaceful* and *Nostalgic* remained statistically significant. The results suggest that VNE-specific ratings did not always align with participants' broader associations with the Kansei terms, highlighting the importance of careful vocabulary selection and validation in multilingual evaluation tools.

5 DISCUSSION

This study explored participants' interpretations of the preliminary VN vocabulary. The discussion reviews the main findings, limitations, and future research directions.

5.1 Word Associations Across the Languages

Luonto / 自然 / *Nature:* Across languages, associations with Nature most commonly referenced greenery, forests, trees, and plants. Silence and peace were frequently mentioned experiential qualities, while home, life, animals, and the environment also appeared, showing that nature was understood as both a physical and emotionally grounding domain. *Outous* / 畏怖 / *Strange:* Associations with Strange centred on fear and curiosity, reflected in repeated terms such as scary and interesting. Words like strange, nature, and animals suggest that unfamiliar or ambiguous environments, often tied to natural elements, were experienced as attention-capturing or unsettling. *Rauha* / 穏やか / *Peaceful:* Responses emphasised tranquillity, natural elements (sky, wind, snow, rain, forest), and states of mental calmness. Feelings of safety and home were also prominent, indicating that peacefulness combined environmental serenity with emotional security and groundedness. *Infakstuuri* / 街並み / *City:* Associations with City focused on built structures (homes, buildings, streets) and urban features such as traffic, cars, and noise. Social themes: people, crowds, homelessness, highlighted human dynamics within cities. Everyday contrasts such as good coffee, history, and Tokyo illustrated diverse urban experiences. *Nostalgia* / 郷想 / *Nostalgic:* Nostalgic showed the most consistent meaning across languages. Associations centred heavily on childhood memories, including play, TV shows, music, films, Gameboy, and photographs. Warmth, home, and summers at the cottage highlighted positive memories. These patterns align with research showing that memories are multisensory (Murray et al., 2004) and that scent can evoke particularly vivid recollections (Verbeek & van Campen, 2013). *Henkisyys* / 神秘 / *Spirituality:* Associations with Spirituality revolved around beliefs, experiences, and personal narratives. Nature, soul, religion, and meditation were recurring terms, suggesting that spirituality was linked to meaning-making, introspection, and transcendent or natural elements. *Ympäristö* / 環境 / *Environment:* Responses were dominated by nature-related words and everyday links to environmental contexts. Participants referenced a wide range of settings: city, streets, trees, soil, sea, forest, alongside broader societal and ecological terms such as climate, pollution, and planet, indicating awareness of environmental and global issues. *Aistillisuus* / きれい / *Beautiful:* Associations with Beautiful centred on aesthetic and emotional qualities, including sensitivity, nature, and expressions such as "something that moves the heart". Beauty was perceived through both affective resonance and natural imagery. *Ilmapiiri* / 雰囲気 / *Atmosphere:* Participants associated Atmosphere with mood and experiential ambience, using terms such as feeling and vibe. Air-related and spatial terms (air, space, sky) suggested that atmosphere was understood through both emotional tone and physical qualities of openness. *Pakoarjesta* / 脱日常 / *Escapism:* Associations highlighted nature, silence, and tranquillity as forms of retreat. Leisure and entertainment, fantasy, TV series, games, and adventure were common pathways for escape, while terms like doomscrolling and escape from reality reflected psychological withdrawal and coping. Overall, the associations illustrate that emotional and semantic interpretations of VN vocabulary are shaped by both shared human experiences and

language-specific cultural frames, underscoring the need for multilingual validation in affective terminology development.

5.2 Emotional Salience vs. Descriptive Accuracy

The results indicate that emotional connection does not consistently align with descriptiveness in VN terminology. Word sets such as Peaceful and Nostalgic showed significant differences between emotional connection ratings and descriptive ratings. This suggests that participants' broader associations with these words may not always align with how well they describe the VNE, highlighting the need for measurement tools that capture both dimensions. Previous cross-linguistic analyses have shown that emotional meanings can vary substantially across languages, reflecting differences in how emotions are categorised and expressed (Sadiqzade, 2025). While the present study did not directly test linguistic differences, these findings highlight the importance of validating multilingual Kansei vocabularies across language contexts. When developing scales for emotional perception of VN, incorporating items that represent emotional impact alongside descriptive clarity may provide a more accurate representation of user experience.

5.3 Cultural Variability in Affective Interpretation

Participants' discussions highlighted that the interpretive logic behind the original vocabulary did not readily transfer to new groups: terms such as *Aistillisuus* / きれい / *Beautiful* were not recognised as parallel in meaning. These reactions reflect broader challenges across languages in representing subjective experience. As prior research shows, emotional perception and expression are shaped by cultural norms, leading to systematic differences in how meanings are interpreted (Scherer et al., 2011). People often rely on intersubjective expectations, beliefs about how others typically feel, rather than on their own preferences (Chiu et al., 2010), and attitudes themselves vary across cultures in how personally or socially anchored they are (Riemer et al., 2014). Consequently, these insights reinforce the importance of carefully selecting and validating multilingual Kansei vocabularies and support the development of a Kansei evaluation tool that addresses variation in affective interpretation in the VN context.

5.4 Study Limitations and Future Work

Although limited by a modest sample size, this exploratory and participatory study provides useful insights for developing tools to evaluate nature experiences. As the descriptiveness and emotional connection questions differed in both focus and context, observed differences should be interpreted cautiously. The study was also not designed to directly test linguistic or cultural differences. Future work will refine and validate the Kansei vocabulary and develop a culturally responsive Kansei Evaluation Tool for use in immersive VR and real-world nature settings.

6 CONCLUSION

This study showed that people from different language backgrounds interpret Kansei words in the VN context through both shared experiences and individual variation. Differences were observed between how certain words described the VNE and participants' broader emotional

associations with them. These findings highlight the importance of carefully selecting and validating multilingual Kansei vocabularies for evaluation tools. Although exploratory, the study raises important questions about multilingual Kansei scale design and provides direction for developing a culturally responsive Kansei Evaluation Tool. Future work will refine the vocabulary and test it in immersive VR and real-world environments to better capture emotional and sensory responses to VNE.

Acknowledgments

This work was supported by Business Finland (AWE project) and the European Regional Development Fund (ProtoXR project).

References

Calogiuri, G., Petersen, E., Haile, A., Flaten, O. E., Fröhlich, F., & Litleskare, S. (2023). The impact of visualization techniques of immersive virtual scenarios in promoting nature connectedness: A blind randomized controlled trial with mixed-methods approach. Journal of Environmental Psychology, 90, 102102.

Capaldi CA, Dopko RL and Zelenski JM. (2014) The relationship between nature connectedness and happiness: a meta-analysis. Front. Psychol. 5:976.

Chiu, C. Y., Gelfand, M. J., Yamagishi, T., Shteynberg, G., & Wan, C. (2010). Intersubjective culture: The role of intersubjective perceptions in cross-cultural research. *Perspectives on Psychological Science*, *5*(4), 482-493.

Harada, A., On the definition of Kansei. In Modeling the Evaluation Structure of Kansei 1998 Conference, volume 2, page 22, 1998.

Kaplan, R., & Kaplan, S. (1989). The experience of nature: A psychological perspective. Cambridge university press.

Larson, L. R., Szczytko, R., Bowers, E. P., Stephens, L. E., Stevenson, K. T., & Floyd, M. F. (2019). Outdoor time, screen time, and connection to nature: troubling trends among rural youth?. Environment and Behavior, 51(8), 966-991.

Lee, S., Harada, A., & Stappers, P. J. (2002). Pleasure with products: Design based on Kansei. *Pleasure with products: Beyond usability*, *219*, 229.

Lévy, P. D. (2013). Beyond kansei engineering: The emancipation of kansei design. *International journal of design*, *7*(2), 83-94.

Lokman, A. M. (2010). Design & emotion: The kansei engineering methodology. *Malaysian Journal of Computing (MJOC)*, *1*(1), 1-14.

Lucero, A. (2015). Using affinity diagrams to evaluate interactive prototypes. In Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15. Springer, 231–248.

Murray, M. M., Michel, C. M., De Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, *21*(1), 125-135.

Nagamachi, M. (2011). Kansei/affective engineering and history of Kansei/affective engineering in the world. *Kansei/affective engineering*, *13*, 1-12.

Nurmela, P., Etto, J., Colley, A., & Häkkilä, J. (2025a). Barefoot Tactile Virtual Autumn Forest Experience. In Proceedings of the 28th International Academic Mindtrek Conference, 247-255.

Nurmela, P., Qureshi, A., & Häkkilä, J. (2025b). The Meaningfulness of Place: Evaluating Nature Postcards from Finland, Japan and Generative AI. In IASDR 2025 International Association of Societies of Design Research: Design next!. Design Research Society.

Riemer, H., Shavitt, S., Koo, M., & Markus, H. R. (2014). Preferences don't have to be personal: Expanding attitude theorizing with a cross-cultural perspective. *Psychological review*, *121*(4), 619.

Sadiqzade, Z. (2025). The Linguistic Expression of Emotion: A Cross-Cultural Analysis. *EuroGlobal Journal of Linguistics and Language Education*, *2*(3), 42-54.

Saeki, J. (2001). Longing for 'Beauty'. In Marra, M. F. (2001). *A history of modern Japanese aesthetics* (Vol. 57). University of Hawaii Press., 25-42.

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401-435.

Schütte, S., Lokman, A. M., Marco-Almagro, L., Ishihara, S., Yanagisawa, H., Yamanaka, T., ... & Coleman, S. (2024). Kansei for the digital era. *International Journal of Affective Engineering*, *23*(1), 1-18.

Ulrich, R.S., Simons, R.F., Losito, B.D., Fiorito, E., Miles, M.A. and Zelson, M., 1991. Stress recovery during exposure to natural and urban environments. Journal of Environmental Psychology, 11(3), pp.201–230.

Verbeek, C., & Van Campen, C. (2013). Inhaling memories: Smell and taste memories in art, science, and practice. *The Senses and Society*, *8*(2), 133-148.

CROSS-CULTURAL RESPONSES TO ADVERTISEMENT DESIGN STYLES
A STUDY OF SUBJECTIVE EVALUATION AND VISUAL ATTENTION

Liwen Zhang[1a], Beibei Feng[a] ,Toru Nakata[2a] and Toshikazu Kato[3a]

[a] *Chuo University, Tokyo, Janpan, a22.5s74@g.chuo-u.ac.jp*

ABSTRACT

This study explores cross-cultural affective responses to culturally inspired advertisement design styles by integrating subjective evaluation and eye-tracking measures. Chinese and Japanese participants evaluated AI-generated product advertisements designed in three Kansei-oriented styles: Chinese, Japanese, and German. The stimuli covered five product categories. A total of 60 participants completed a Kansei evaluation questionnaire, and a subset of 30 participants underwent eye-tracking using a Tobii Pro Spark. The results showed that advertisement style had a clearer influence on affective evaluation than participant nationality. Chinese-style advertisements elicited higher perceived self-fit and purchase/use intention, whereas German-style advertisements evoked higher trustworthiness; Japanese-style advertisements tended to receive lower ratings. Participant nationality showed relatively limited effects on subjective evaluation. Eye-tracking results showed that visit count produced the clearest nationality-related pattern, with Chinese participants revisiting advertisements more frequently than Japanese participants. Visit count showed a moderate positive association with fixation duration, but little association with subjective evaluation items, suggesting that repeated viewing may reflect visual engagement, comparison, or confirmation rather than direct positive evaluation. These findings demonstrate the value of combining subjective responses and eye-tracking measures to understand cross-cultural affective responses to advertisement design.

Keywords: *Cross-culture, Affective design; Advertisement; Eye tracking; Visual attention*

[1]1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. Email: a22.5s74@g.chuo-u.ac.jp.

[2]1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. Email: toru-nakata.77b@g.chuo-u.ac.jp

[3]1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. Email: toshi.vader.k.08r@g.chuo-u.ac.jp

.

1 INTRODUCTION

In globalized consumer markets, advertisements are increasingly designed for audiences from different cultural backgrounds. Advertisement design does not only communicate product information but also conveys culturally embedded meanings through visual style and persuasive cues.

Previous studies have shown that advertising appeals matched with consumers' cultural backgrounds and product-use conditions can influence advertising effectiveness [1]. Therefore, understanding how culturally embedded visual cues evoke specific emotional and Kansei responses is crucial for both cross-cultural marketing and Kansei engineering.

Previous studies on cross-cultural advertising have often examined advertising effectiveness through self-reported evaluations, such as attitude toward the advertisement, perceived trustworthiness, and purchase intention. Research on ad-self congruency also suggests that the perceived fit between an advertisement and the consumer's self-concept can affect ad evaluation and purchase-related responses [2].

However, subjective evaluation alone cannot fully reveal consumers' visual processing during advertisement viewing.

To address this issue, the present study combines questionnaire evaluation and eye-tracking data to examine cross-cultural responses to advertisement design styles. Since visual attention to advertisement elements plays an important role in advertisement processing [4], integrating subjective evaluation with eye-tracking measures can provide a more comprehensive understanding of how advertisement style influences consumer responses.

Specifically, this study compares Chinese and Japanese participants' responses to three advertisement design styles: Chinese-style, Japanese-style, and German-style advertisements. The advertisement stimuli were generated as virtual product posters based on reference advertisements from each country, allowing culturally different design styles to be compared under the same product categories.

Five product categories were used as stimuli. Subjective evaluations, including ease of 1) understanding, 2) naturalness, 3) attractiveness, 4) trustworthiness, 5) self-fit, and 6) purchase/use intention, were collected using a 5-point scale, and visual attention was measured using Tobii eye-tracking indicators, including time to first fixation, fixation duration, and visit count.

The purpose of this study is to examine whether subjective evaluation and visual attention show consistent or different patterns, and whether cultural differences emerge differently across stages of visual attention.

2 RELATED WORK AND RESEARCH AIM

Cross-cultural advertising research has shown that consumers' responses to advertisements are influenced by cultural values, communication styles, and familiar visual expressions. Advertisements congruent with consumers' cultural backgrounds or product-use contexts may be evaluated more positively because they are easier to understand and more consistent with consumers' expectations [1]. However, advertising effectiveness cannot be explained only by national or cultural congruency. Research on ad-self congruency suggests that consumers may respond more favorably when an advertisement feels personally relevant or fits their self-concept [2].

In Kansei engineering, culturally meaningful visual styles can be regarded as design stimuli that evoke psychological and emotional responses [3]. Therefore, advertisement style may influence multiple dimensions of consumer evaluation, such as self-fit, trustworthiness, and purchase/use intention. However, self-reported evaluation alone cannot fully reveal how consumers visually

process advertisements. Eye-tracking can complement questionnaire data by capturing visual attention during advertisement viewing. Previous advertising research has used eye-tracking to examine how attention is distributed across advertisement elements, such as images, text, and brand information [4].

Based on this background, the present study explores how Chinese and Japanese participants respond to Chinese-, Japanese-, and German-style advertisements across subjective evaluation and eye-tracking indicators. Because advertisement style involves multiple visual and semantic elements, this study treats it as a composite design condition rather than an isolated single-variable manipulation.

3 METHOD

3.1 Participants

A total of 60 participants completed the questionnaire survey. Thirty participants were Chinese and thirty participants were Japanese. Participants were between 20 and 40 years old. Among them, 30 participants also took part in the eye-tracking experiment, including 15 Chinese participants and 15 Japanese participants. All participants had normal or corrected-to-normal vision

3.2 Stimuli

The stimuli consisted of advertisements for five product categories: beverages, mattresses, earrings, computers, and game consoles. For each product category, three advertisement design styles were prepared: Chinese-style, Japanese-style, and German-style advertisements. Thus, a total of 15 advertisement stimuli were used in the experiment.

The advertisements were created as virtual product posters rather than real commercial advertisements. Reference posters from China, Japan, and Germany were collected through online searches and used to identify representative visual characteristics, such as layout, color, information density, imagery, typography, and textual expression. Based on these references, a text-to-image generative AI tool was used to generate posters for fictional products. The generated posters were then screened and selected by the researchers to ensure that each stimulus reflected the intended style while avoiding real brands, existing products, or identifiable commercial materials.

To reduce the “black-box” nature of AI-generated stimuli, the main visual characteristics of the three styles were summarized after generation and screening, as shown in Table 1.



Advertisement style was treated as a composite design condition rather than an isolated manipulation of a single visual variable. Therefore, the terms Chinese-style, Japanese-style, and German-style refer to visual advertising style cues generated and selected for this experiment, rather than direct representations of national cultures as a whole.

Figure 1. Example of AI-generated advertisement stimuli used in the experiment. The advertisements were generated based on reference posters from each country and were used as culturally inspired design style conditions.

Table 1. Summary of visual characteristics of the three advertisement styles.

Style	Main visual characteristics	Design interpretation
Chinese	Relatively high information density; vivid colors; compact layout; stronger decorative or promotional elements	Emphasis on visual richness, promotional impact, and expressive presentation
Japanese	Higher whitespace; softer color tones; restrained composition; balanced arrangement of image and text	Emphasis on subtlety, balance, calmness, and gentle impression
German	Minimalist composition; stronger contrast; clearer grid-like layout; product-centered presentation; functional typography	Emphasis on clarity, functionality, order, and reliability

3.3 Procedure

In each trial, three advertisements from the same product category were presented simultaneously on the screen. The presentation position was fixed: the Japanese-style advertisement was placed on the left, the Chinese-style advertisement in the center, and the German-style advertisement on the right. The stimulus order was also fixed across participants.

For the questionnaire survey, participants evaluated the advertisements using a 5-point Likert scale. For the eye-tracking experiment, participants were seated in front of the display, and their eye movements were recorded using a Tobii Pro Spark eye tracker. The eye tracker was calibrated before the task. The Tobii Pro Spark records gaze data at 60 Hz under its default setting. Each advertisement image was defined as an area of interest.

After viewing the advertisements, participants completed subjective evaluations.

3.4 Measure

3.4.1 Questionnaire Measures

Subjective evaluations were collected using a 5-point scale. The questionnaire consisted of six items related to understandability, naturalness, attractiveness, trustworthiness, self-fit, and purchase/use intention, as shown in Table 2. Higher scores indicated more positive evaluations.

Table 2. **Questionnaire items used for subjective evaluation**

Construct / Item	Description
Ease of understanding	The advertisement is easy to understand.
Naturalness	The advertisement feels natural and has little sense of incongruity.

Attractiveness	The advertisement is attractive.
Trustworthiness	The advertisement appears trustworthy.
Self-fit	The advertisement feels targeted toward me / domestic consumers.
Purchase/use intention	I would like to purchase or use the product.

3.4.2 Eye-Tracking Measures

Eye movements were recorded using a Tobii Pro Spark eye tracker. Three eye-tracking indicators were analyzed: time to first fixation, fixation duration, and visit count. Time to first fixation was used as an indicator of initial attention capture. Fixation duration was used as an indicator of visual engagement or information processing. Visit count was used as an indicator of repeated viewing or revisiting behavior.

3.5 Data Analysis

For the questionnaire data, mean scores were calculated for each participant, advertisement style, and evaluation item. The six questionnaire items were analyzed separately.

For the eye-tracking data, time to first fixation, fixation duration, and visit count were analyzed for each advertisement style. Data were averaged across the five product categories for each participant before statistical analysis. Mixed-design analyses of variance were conducted with participant nationality as a between-subject factor and advertisement style as a within-subject factor. When significant main effects or interactions were found, post-hoc pairwise comparisons were conducted. Statistical significance was set at $p < .05$.

4 RESULTS

4.1 Subjective Evaluation

Subjective evaluation ratings were analyzed for six items: 1) understandability, 2) naturalness, 3) attractiveness, 4) trustworthiness, 5) self-fit, and 6) purchase/use intention. The following analysis focuses on three items most directly related to affective advertising effectiveness: 5) self-fit, 6) purchase/use intention, and 4) trustworthiness. Understandability, naturalness, and attractiveness were treated as supplementary items for checking the general impression and basic quality of the stimuli.

Fig. 2 shows boxplots of participant-level mean ratings for self-fit, purchase/use intention, and trustworthiness. For each measure, ratings were averaged across the five product categories for each participant and advertisement style before visualization. Participant nationality was included as a between-subject factor in the statistical analyses, but ratings were combined across nationality groups in Fig. 2 to show the overall advertisement-style pattern.

JP-style advertisements tended to receive lower ratings across the three representative items. CN-style advertisements showed relatively higher ratings for self-fit and purchase/use intention, whereas DE-style advertisements showed the highest ratings for trustworthiness.

For self-fit, a mixed-design ANOVA showed a significant main effect of advertisement style, $F(2,116) = 15.47$, $p < .001$, $\eta p^2 = .211$. The main effect of nationality and the interaction were not significant. Bonferroni-corrected pairwise comparisons showed that CN-style and DE-style advertisements were rated significantly higher than JP-style advertisements, while CN-style and DE-style advertisements did not differ significantly.

For purchase/use intention, significant main effects were found for nationality, $F(1,58) = 5.50$, $p = .022$, $\eta p^2 = .087$, and advertisement style, $F(2,116) = 19.98$, $p < .001$, $\eta p^2 = .256$. Japanese

participants gave higher purchase/use intention ratings overall than Chinese participants. Pairwise comparisons showed that CN-style and DE-style advertisements were rated significantly higher than JP-style advertisements, with no significant difference between CN-style and DE-style advertisements. The interaction was not significant, $F(2,116) = 2.78$, $p = .066$.

For trustworthiness, advertisement style also showed a significant main effect, $F(2,116) = 29.90$, $p < .001$, $\eta p^2 = .340$. The main effect of nationality and the interaction were not significant. DE-style advertisements received the highest trustworthiness ratings, followed by CN-style and JP-style advertisements. Post-hoc comparisons indicated that DE-style advertisements were rated significantly higher than JP-style advertisements in both nationality groups.

Overall, the subjective evaluation results indicate that advertisement style had a clearer effect than participant nationality. JP-style advertisements were evaluated less positively, whereas CN-style and DE-style advertisements received higher evaluations in different dimensions.

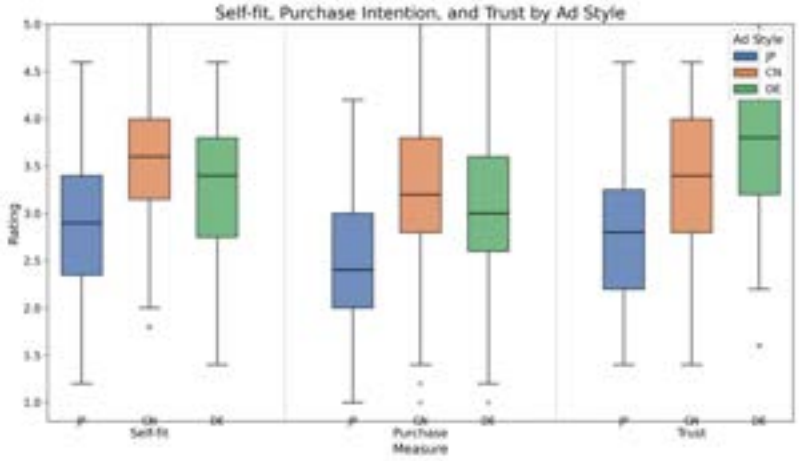


Fig. 1. Boxplots of self-fit, purchase/use intention, and trustworthiness by advertisement style. For each measure, participant-level mean ratings are shown for JP-, CN-, and DE-style advertisements.

4.2 Subjective Evaluation by Participant Nationalit

To examine nationality-related tendencies in subjective evaluation, the six questionnaire items were compared between Chinese and Japanese participants. Fig. 3 shows boxplots of participant-level mean ratings for understandability, naturalness, attractiveness, trustworthiness, self-fit, and purchase/use intention. Ratings were averaged across advertisement styles and product categories for each participant before visualization.

Japanese participants showed slightly higher median ratings for several items, particularly attractiveness, self-fit, and purchase/use intention. However, the distributions of the two nationality groups largely overlapped across the six items.

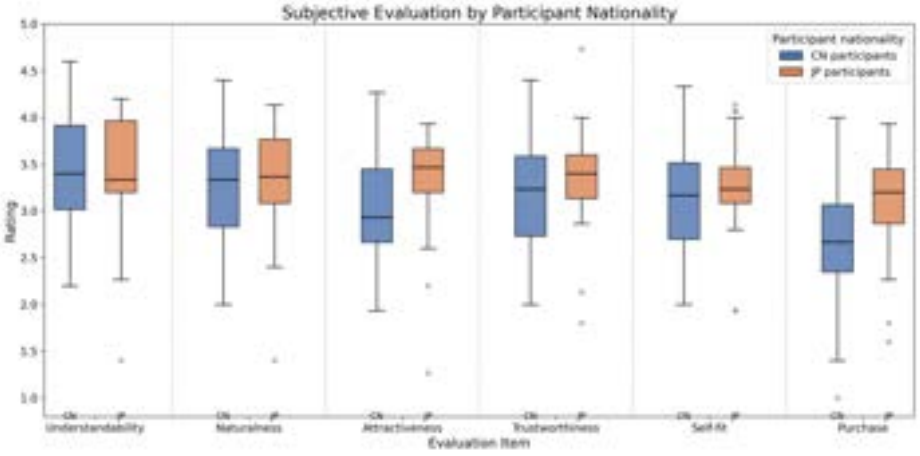


Fig. 2. Boxplots of six subjective evaluation items by participant nationality. Each box represents the distribution of participant-level mean ratings for Chinese and Japanese participants. Ratings were averaged across advertisement styles and product categories for each participant before visualization.

4.3 Eye-Tracking Measures

Eye-tracking data were analyzed using three indicators: time to first fixation (TTFF), fixation duration, and visit count. TTFF was used to examine initial attention capture, fixation duration was used to examine visual engagement, and visit count was used to examine repeated viewing behavior.

Because the presentation position of each advertisement style was fixed across trials, TTFF was not interpreted as a direct indicator of advertisement style preference. Instead, it was treated as a descriptive baseline reflecting initial viewing tendency under the fixed layout.

TTFF showed a significant main effect of advertisement style and a significant interaction with nationality. However, because advertisement style and screen position were fixed, TTFF was interpreted only as a style/position-related baseline rather than as evidence of advertisement-style preference.

For fixation duration, a significant main effect of advertisement style was found ($F(2,56) = 3.90$, $p = .026$). The main effect of participant nationality was not significant ($F(1,28) = 1.66$, $p = .209$), and the interaction did not reach statistical significance, ($F(2,56) = 2.89$, $p = .064$). Post-hoc comparisons suggested that the nationality difference was significant only for JP-style advertisements, indicating that cultural differences in viewing duration were limited.

Visit count showed the clearest nationality-related pattern. Significant main effects were found for participant nationality($F(1,28) = 21.90$, $p < .001$), and advertisement style ($F(2,56) = 32.38$, $p < .001$). The interaction was also significant, ($F(2,56) = 4.51$, $p = .015$). As shown in Fig. 4, Chinese participants tended to show higher visit counts than Japanese participants across all three advertisement styles, although the distributions overlapped to some extent. Post-hoc comparisons confirmed that this nationality difference was significant for CN-, DE-, and JP-style advertisements.

To further examine the meaning of repeated viewing behavior, correlation analyses were conducted between visit count and selected measures, including fixation duration, TTFF, self-fit, purchase/use intention, and trustworthiness. Spearman's rank correlation was used because the eye-tracking sample size was relatively small and visit count was treated as a behavioral indicator of repeated viewing. The correlations were calculated at the condition level based on nationality × product category × advertisement style combinations.

As shown in Table 3, visit count showed a moderate positive association with fixation duration (correlation $\rho = .502$, $p < .005$). This suggests that advertisements revisited more frequently tended to receive longer total visual attention. In contrast, visit count showed little association with TTFF (correlation $\rho = .081$, $p = .675$), or with the selected subjective evaluation items, including self-fit, purchase/use intention, and trustworthiness.

Overall, the eye-tracking results indicate that TTFF was strongly affected by the style/position condition, fixation duration showed limited cultural differences, and visit count showed the clearest nationality-related difference. The correlation analysis further suggests that visit count may be more closely related to visual engagement than to subjective preference or positive evaluation.

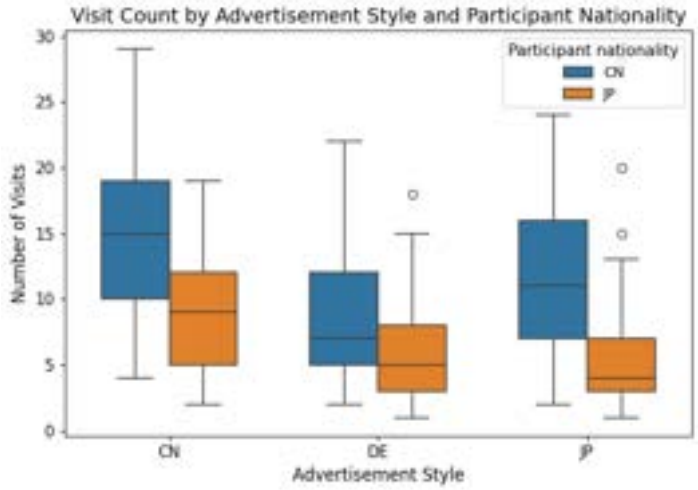


Fig. 3. Boxplots of visit count by advertisement style and participant nationality. Each box represents the distribution of participant-level mean visit counts for Chinese and Japanese participants.

Table 3. Spearman correlations between visit count and selected measures.

Measure	Spearman's ρ	p-value	Significance
Fixation duration	.502	.005	Yes
TTFF	.081	.675	-
Self-fit	.039	.837	-
Purchase/use intention	-.071	.711	-
Trustworthiness	-.130	.495	-

Note. Correlations were calculated at the condition level based on nationality × product category × advertisement style combinations.

5 DISCUSSION

5.1 Subjective Evaluation and Advertisement Style

The subjective evaluations reveal a distinct style-driven rather than nationality-driven pattern. The association of DE-style advertisements with trustworthiness and CN-style advertisements with self-fit and purchase/use intention suggests that culturally inspired visual cues may trigger different affective dimensions of advertising evaluation.

CN-style advertisements may have triggered stronger emotional resonance, familiarity, and perceived approachability. This indicates that visually familiar or emotionally accessible styles may be effective in evoking positive Kansei connections with consumers' self-concept. Rather

than simply reflecting a preference for one's own cultural background, this result suggests that certain visual styles may enhance perceived relevance and emotional closeness.

DE-style advertisements, on the other hand, appeared to evoke stronger cognitive and affective trust. One possible explanation is the country-of-origin effect [5]. Country-related visual cues may activate consumers' existing associations with particular countries, even when the products themselves are fictional and not actually produced in those countries. In this study, DE-style advertisements may have activated associations such as precision, functionality, reliability, and technological sophistication. These associations may have contributed to the higher trustworthiness ratings observed for DE-style advertisements.

Overall, the results do not support a simple interpretation that consumers always prefer advertisements from their own cultural background. Instead, advertisement style appears to influence evaluation through multiple Kansei dimensions, such as perceived relevance, emotional resonance, purchase appeal, and trust. In this sense, culturally inspired advertisement styles should be understood not only as visual aesthetic variations, but also as cues that may activate broader product- and country-related associations.

5.2 Participant Nationality and Subjective Evaluation

The comparison between Chinese and Japanese participants suggests that participant nationality did not strongly separate subjective evaluation patterns. Although some items showed slightly higher ratings among Japanese participants, the overlap between the two groups indicates that nationality alone was not the dominant factor shaping affective responses.

Rather than showing a straightforward local-style preference, participants from both nationality groups appeared to respond to the affective meanings conveyed by visual style, such as familiarity, self-relevance, trustworthiness, and perceived product quality.

From a Kansei design perspective, this result indicates that cultural style should not be treated as a fixed national label. Rather, culturally inspired visual cues may function as design resources that communicate different affective values. A familiar and expressive style may enhance self-fit and purchase/use intention, whereas a clearer and more functional style may enhance trustworthiness. Therefore, the effect of advertisement style should be interpreted in relation to the intended Kansei value rather than only in relation to the nationality of the target audience.

5.3 Eye-Tracking Results and Visual Attention

The eye-tracking results provided additional insight into how participants visually processed the advertisements. TTFF showed strong differences among advertisement style conditions, but because advertisement style and screen position were fixed, it was interpreted only as a supplementary baseline of initial viewing tendency rather than as an indicator of style preference.

Fixation duration showed a weaker pattern. Although advertisement style had a significant main effect, nationality-related differences were limited and appeared mainly in the JP-style condition. This suggests that total viewing time was not consistently shaped by participant nationality.

In contrast, visit count showed the clearest nationality-related difference, with Chinese participants revisiting advertisements more frequently than Japanese participants across all three styles. This indicates that repeated viewing behavior may be more sensitive to participant background than initial fixation timing or total viewing duration. The correlation analysis further showed that visit count was associated with fixation duration but not with subjective evaluation items, suggesting that frequent revisiting may reflect visual engagement, comparison, or confirmation rather than direct preference, trust, or purchase intention.

This pattern may be related to culturally shaped information-search strategies. Previous studies on holistic and analytic cognitive styles suggest that cultural background can influence how people distribute visual attention across objects, contexts, and relationships [7, 8]. However, because both Chinese and Japanese participants belong to East Asian cultural contexts and cognitive style was not directly measured, this interpretation should be treated as exploratory.

Overall, the eye-tracking results suggest that the clearest nationality-related difference appeared in repeated viewing behavior rather than in subjective evaluation. This supports the value of combining questionnaire data with eye-tracking measures when studying cross-cultural responses to advertisement design.

5.4 Design Implications

These findings suggest that culturally inspired advertisement styles should be selected according to the intended Kansei value rather than the nationality of the target audience. Expressive and familiar styles may support self-fit and purchase/use intention, whereas clear and functional styles may enhance trustworthiness. In addition, repeated viewing should not be interpreted as direct preference, because visit count may reflect comparison or confirmation rather than positive evaluation.

6 LIMITATIONS AND CONCLUSION

6.1 Limitations

This study has several limitations. First, advertisement style was treated as a composite condition including layout, color, imagery, typography, and textual expression; therefore, the effects of individual design elements cannot be isolated. Second, the presentation position of each style was fixed, so TTFF should be interpreted cautiously as a style/position-related measure. Third, the AI-generated stimuli may have reflected simplified or stereotypical representations of national advertising styles, and the eye-tracking sample size was relatively small. Future studies should counterbalance stimulus positions, validate perceived cultural style, manipulate specific design elements, and use larger samples across more product categories.

6.2 Conclusion

This study examined Chinese and Japanese participants' responses to Chinese-, Japanese-, and German-style advertisements by combining subjective evaluation and eye-tracking data. Advertisement style showed a clearer influence on subjective evaluation than participant nationality. CN-style advertisements were associated with higher self-fit and purchase/use intention, whereas DE-style advertisements were associated with higher trustworthiness; JP-style advertisements tended to receive lower ratings. Eye-tracking results showed that visit count, rather than TTFF or fixation duration, revealed the clearest nationality-related pattern, with Chinese participants revisiting advertisements more frequently. Because visit count was associated with fixation duration but not with subjective evaluation, repeated viewing may reflect visual engagement, comparison, or confirmation rather than direct preference. These findings demonstrate the value of combining subjective and eye-tracking measures to understand cross-cultural Kansei responses to advertisement design styles.

ACKNOWLEDGMENTS

The author gratefully acknowledges Prof. Patrick Rau at Tsinghua University for his valuable support and assistance during this study.

REFERENCES

1. Zhang, Y., Gelb, B.D.: Matching advertising appeals to culture: The influence of products' use conditions. Journal of Advertising 25(3), 29–46 (1996). https://doi.org/10.1080/00913367.1996.10673505
2. Chang, C.: Self-congruency as a cue in different advertising-processing contexts. Communication Research 29(5), 503–536 (2002). https://doi.org/10.1177/009365002236193

3. Nagamachi, M.: Kansei Engineering: A new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics 15(1), 3–11 (1995). https://doi.org/10.1016/0169-8141(94)00052-5
4. Pieters, R., Wedel, M.: Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. Journal of Marketing 68(2), 36–50 (2004). https://doi.org/10.1509/jmkg.68.2.36.27794
5. Al-Sulaiti, K.I., Baker, M.J.: Country of origin effects: A literature review. Marketing Intelligence & Planning 16(3), 150–199 (1998). https://doi.org/10.1108/02634509810217309
6. Roth, M.S., Romeo, J.B.: Matching product category and country image perceptions: A framework for managing country-of-origin effects. Journal of International Business Studies 23(3), 477–497 (1992). https://doi.org/10.1057/palgrave.jibs.8490276
7. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: Holistic versus analytic cognition. Psychological Review 108(2), 291–310 (2001). https://doi.org/10.1037/0033-295X.108.2.291
8. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences 102(35), 12629–12633 (2005). https://doi.org/10.1073/pnas.0506162102

DECODING THE OLFACTORY CHARM: A MIRYOKU ENGINEERING AND P-A MAPPING ANALYSIS OF WESTERN FRAGRANCES IN EAST ASIA

Jia-Huei Yu, Min-Yuan Ma, Tseng-Ping Chiu

Department of Industrial Design, National Cheng Kung University, Taiwan,
p36144128@gs.ncku.edu.tw, mamy@mail.ncku.edu.tw, mattchiu@gs.ncku.edu.tw

ABSTRACT

As the global fragrance industry expands, Western fragrance design often overlooks East Asian emotional preferences. This study applies Miryoku Engineering to identify the psychological drivers of scent attractiveness among Taiwanese consumers. The Evaluation Grid Method (EGM) analyzed 44 fragrance brands, followed by a questionnaire with 62 experienced users. Results showed that fragrance is mainly used in private or work spaces for relaxation (70.3%) and spatial transformation (52.7%), with woody scents (56.8%) being most preferred. Quantification Theory Type I and the Pleasure–Arousal (P-A) model revealed that Relaxation had the highest explanatory power (R = 0.524), driven by moist moss notes and label–scent consistency. These findings indicate a preference for high-pleasure, low-arousal experiences and provide a Kansei design framework for developing nature-inspired fragrances for smart homes and wellness environments.

Keywords: Kansei Engineering, Miryoku Engineering, Olfactory Attractiveness, Cross-Cultural Affect

1 INTRODUCTION

In the contemporary landscape of sensory marketing, fragrance products have evolved beyond their functional role to become important media for emotional regulation and the creation of domestic spatial experiences. As the global fragrance market continues to expand, Western fragrance brands have become dominant in product development and branding strategies. These products often emphasize individualistic expression, novelty, and high-arousal sensory stimulation. In contrast, East Asian consumers tend to value harmony, restraint, and emotional balance, suggesting that fragrance perception is shaped not only by physiological responses but also by cultural values and lived experiences. Previous studies have demonstrated that the familiarity and pleasantness of scents vary according to environmental, seasonal, and cultural contexts (Seo et al., 2009). Likewise, Ideal Affect Theory proposes that Eastern cultures generally prefer low-arousal positive emotions, whereas Western cultures favor high-arousal positive emotions (Tsai, 2007). These cultural differences imply that fragrance attractiveness should be understood within its cultural context rather than through universal sensory principles alone.

Previous fragrance research has primarily focused on chemical composition, odor classification, and physiological or emotional responses to scent (Thiboud, 1991; Kiecolt-Glaser et al., 2008).

Although these studies explain how fragrances influence human perception and mood, they provide limited guidance for translating users' subjective emotional preferences into practical product design strategies. Within Kansei Engineering, existing studies have successfully quantified relationships between product attributes and emotional responses across various product categories. However, relatively few studies have investigated how fragrance attractiveness can be systematically transformed into actionable design attributes under different cultural contexts. In particular, research integrating cross-cultural affective theory with fragrance-oriented Kansei Engineering remains limited.

To address this gap, this study adopts Miryoku Engineering, an extension of Kansei Engineering that focuses on revealing the hierarchical cognitive structure underlying product attractiveness. Unlike conventional Kansei Engineering, which primarily establishes direct relationships between product attributes and Kansei responses, Miryoku Engineering employs the Evaluation Grid Method (EGM) to identify the psychological hierarchy linking abstract emotional values, evaluation criteria, and concrete design attributes (Sanui, 1996; Ujigawa, 1997; Nagamachi, 2002). This hierarchical approach is particularly suitable for fragrance design because olfactory attractiveness is highly subjective, experience-dependent, and culturally constructed. By uncovering why users perceive certain fragrances as attractive, Miryoku Engineering provides richer design knowledge than emotion mapping alone.

Accordingly, this study investigates the olfactory attractiveness of Western fragrance candles among Taiwanese consumers by integrating Miryoku Engineering with the Pleasure–Arousal (P–A) model (Russell, 1980). First, the Evaluation Grid Method is employed to construct the hierarchical structure of fragrance attractiveness. Subsequently, Quantification Theory Type I is applied to identify the concrete fragrance attributes that contribute to different affective dimensions, and the results are mapped onto the Pleasure–Arousal space to examine their cultural affective characteristics. The findings contribute to Kansei Engineering by providing a systematic framework for translating culturally specific fragrance preferences into actionable design guidelines, thereby supporting fragrance product development, wellness-oriented environments, and cross-cultural sensory design.

2 LITERATURE REVIEW

2.1 Kansei Engineering and Miryoku Engineering

Kansei Engineering is a human-centered design methodology that translates users' emotional responses into product design elements, enabling designers to systematically connect subjective perceptions with objective design attributes (Nagamachi, 2002). It has been widely applied in product design, packaging, architecture, and consumer products to establish quantitative relationships between product features and users' affective evaluations.

However, conventional Kansei Engineering primarily focuses on identifying direct relationships between design attributes and emotional responses, offering limited insight into the cognitive process through which users construct perceptions of product attractiveness. To overcome this limitation, Miryoku Engineering was proposed as an extension of Kansei Engineering that

emphasizes the psychological structure of "Miryoku" (attractiveness or charm) rather than emotional evaluation alone (Ujigawa, 1997). Its core tool, the Evaluation Grid Method (EGM), organizes user evaluations into three levels through laddering interviews: upper-level (affective values), middle-level (evaluation items), and lower-level (product attributes). This hierarchical approach reveals not only what users prefer but also why specific attributes evoke particular emotions, making it especially suitable for culturally influenced and experience-based products such as fragrances.

2.2 The Relationship Between Olfaction and Emotion

Scents directly influence the limbic system, including the amygdala and hippocampus, thereby affecting emotions, moods, and physiological responses (Kiecolt-Glaser et al., 2008). Chiu (2010) further demonstrated that fragrance characteristics are closely associated with affective responses, indicating that different scent attributes evoke distinct emotional experiences during product use.

Based on these findings, Kansei Engineering has been applied to fragrance research to examine the relationships between scent characteristics and emotional perceptions, including pleasantness, relaxation, freshness, and the Pleasure–Arousal (P–A) dimensions (Chiu, 2010). However, most studies focus on evaluating predefined fragrance attributes, while limited attention has been given to the hierarchical construction of fragrance attractiveness and the influence of cultural differences. This limitation highlights the need to integrate cross-cultural perspectives into fragrance design, which is discussed in the following section.

2.3 Ideal Affect Theory and Cross-Cultural Differences

The Circumplex Model of Affect, proposed by Russell (1980), is the cornerstone of contemporary emotional psychology. This model defines human emotions along two orthogonal axes:

1. Pleasure (P-axis): The degree of positive or negative valence.
2. Arousal (A-axis): The degree of physiological and psychological activation.

Stanford Professor Jeanne Tsai's "Ideal Affect Theory" indicates that different cultures have distinct preferences for the "ideal position" within these emotional coordinates. Western individualistic cultures tend to favor "High Arousal Positive (HAP)" states, such as excitement and enthusiasm, corresponding to the first quadrant of the Pleasure-Arousal (P-A) matrix. In contrast, East Asian collectivistic cultures prefer "Low Arousal Positive (LAP)" states, such as calmness and harmony, corresponding to the fourth quadrant. This study hypothesizes that the perception of "charm" in fragrance products among East Asian audiences will shift toward the low-arousal quadrant.

3 RESEARCH METHODOLOGY

3.1 Conversion of Secondary Data into EGM

This study conducted text analysis and netnographic observation using scented candle semantic data shared by two prominent East Asian fragrance YouTubers, Perfume Fanatic (香水有毒) and Sherry (雪莉波波貓). A focus group of three experts in Kansei Engineering and industrial design translated the semantic descriptions into the Evaluation Grid Method (EGM) hierarchy. Through consensus, similar expressions were merged, duplicates removed, and descriptors classified into upper-level affective values, middle-level evaluation items, and lower-level fragrance attributes, establishing a standardized and reliable EGM framework for subsequent quantitative analysis.

The data corpus consists of 20 videos from each expert, released on their respective platforms between 2020 and 2025. Collectively, 44 Western fragrance candle brands were identified and analyzed.

Table 1. Statistics of Fragrance Candle Brands

Country	Number of brands	Name of brands
America	13	VOLUSPA, PADDYWAX, AYDRY. CO, Bath and body works, P.F. CANDLE, Aspen Bay Candle, Brooklyn Candle Studio, NEST, MALIN+GOETZ, Aquiesse, Tocca, Boy Smells, Nomad Noe
France	8	Diptyque, CIRE TRUDON, carrière freres, Atelier Cologne, Maison Louis Marie, MAISON MARGIELA, MAISON FRANCIS KURKDJIAN, Éditions de Parfums Frédéric Malle
Australia	7	Wavertree & London, LUMIRA, CIRCA, GLASSHOUSE, Natio, BULY , Aesop
the UK	7	Jo Malone, True grace, OSKIA, Ted Baker, NEOM, EARL OF EAST, Bamford London
Switzerland	3	IKEA, H & M, Byredo
Italy	2	CULTI MILANO, ACQUA DI PARMA
Spain	1	ZARA
Denmark	1	SKANDINAVISK
Israel	1	Sabon
New Zealand	1	The Aromatherapy Company

The upper, middle, and lower-level levels mentioned by both YouTubers were aggregated, and the terminology was standardized. Based on the original contextual meaning, atmospheric descriptions were categorized as middle-level items (Evaluation Items), while scent descriptions were categorized as lower-level items (Concrete Features). After statistical analysis, five upper-level factors were identified as the core affective dimensions: Freshness, Warmth, Security, Pleasure, and Relaxation.



Figure 1. The EGM Hierarchical Structure of Fragrance Candle Perception.

3.2 Transforming EGM into a Questionnaire

Following the established questionnaire methodology of Miryoku Engineering, the upper-level factors identified from the EGM analysis served as the primary question dimensions. For each section, the corresponding middle-level items were evaluated using a 5-point Likert scale to measure their perceived importance. Additionally, respondents were required to select the most significant lower-level option (concrete feature) associated with each middle-level item.

Concurrently, the questionnaire collected demographic and behavioral data from the participants, including gender, age, scent profile preferences, purchase channels for fragrance candles, usage frequency, and duration of experience.

4 RESULTS

4.1 Data Collection and Descriptive Statistics

To ensure the reliability of the findings within a specific cultural context, a total of 71 questionnaire responses were collected through an anonymous survey distributed to Taiwanese participants. This research was conducted with strict adherence to ethical standards, including obtaining informed consent and ensuring data privacy for all respondents. After filtering for participants with prior experience in using fragrance candles, 62 valid responses were identified. The sample consisted of 42 female and 20 male respondents. In terms of age distribution, 56.3% of the participants were aged 21 to 30, followed by 25.4% aged 31 to 40, indicating that the respondents are predominantly young adults.

Regarding usage context, 62.2% of respondents primarily use fragrance candles in the bedroom, while 37.8% use them in study or workspaces. The primary motivations for use were relaxation (70.3%) and spatial scent transformation (52.7%). In terms of subjective scent preferences, woody notes were the most favored (56.8%), followed by floral notes (50%).

4.2 Analytical Framework: Quantification Theory Type I

The analysis was segmented based on the five upper-level affective appeals. Quantification Theory Type I was employed to identify the relationship between affective perceptions and physical features:

- Dependent Variable (Y): The average preference scores of the middle-level evaluation items.
- Independent Variables (X): The encoded lower-level concrete features (scent characteristics).

By performing this analysis, the study quantifies the impact of specific scent characteristics on the overall "charm" of the product. The results are detailed in the following sections.

Table 2. Relationship between Upper-level Kansei Values and Key Fragrance Attributes

Attractiveness Construct (Upper)	Multiple Correlation (R)	Key Concrete Feature (Lower)	The score of Key Concrete Feature	Corresponding middle-level of the top-scoring category
Relaxation	0.524	Moist Moss Scent	0.395	Forest greenery ambiance
Freshness	0.447	Mossy Scent	0.464	Natural greenery ambiance
Pleasure	0.424	Fig Scent	0.162	Sweet Ambiance
Warmth	0.381	Christmas dessert Scent	0.421	Cozy domestic ambiance
Security	0.301	Scent profiles that faithfully reflect product labeling	0.199	Unpretentious Natural Ambiance

The explanatory power of each model was evaluated using the multiple correlation coefficient (R). The obtained R values ranged from 0.301 to 0.524, suggesting meaningful relationships between the upper-level affective constructs and the encoded lower-level fragrance attributes. Based on the category scores generated by Quantification Theory Type I, the lower-level attribute with the highest score for each upper-level construct was identified as its key design attribute, as summarized in Table 2.

5 DISCUSSION

5.1 Analysis of High Pleasure and High Arousal (HAP) Dimensions

According to Russell's Pleasure–Arousal (P–A) model, Freshness and Pleasure are positioned within the High Pleasure–High Arousal (HAP) quadrant. Within the Miryoku Engineering framework, these dimensions represent affective values that are supported by specific evaluation items and concrete fragrance attributes. The following sections discuss how these hierarchical relationships can be translated into fragrance design knowledge.

5.1.1 Freshness

The Quantification Theory Type I analysis identified “Freshness” as the second strongest affective construct (R = 0.447). The highest-scoring lower-level attribute was “Moss Scent”, which contributed most strongly to the middle-level evaluation item of “Natural Greenery Ambiance”. This hierarchical relationship suggests that Taiwanese consumers interpret freshness not merely as an olfactory sensation but as an environmental experience closely connected with nature.

Previous studies associated freshness with green, citrus, and marine notes (Zarzo, 2013). Müller (1992) further reported regional differences in freshness perception. In contrast, Taiwanese participants associated freshness with moss-related scents, suggesting that freshness is interpreted through ecological familiarity rather than sensory stimulation. These findings suggest that the emotional perception of freshness is shaped by culturally specific cognitive associations, in which the abstract affective value of freshness is supported by the evaluation of natural greenery and ultimately translated into moss-related fragrance attributes.

5.1.2 Pleasure

Pleasure obtained a multiple correlation coefficient of R = 0.424. The highest-scoring lower-level attribute was “Fig Scent”, corresponding to the middle-level evaluation item “Sweet Ambiance”. This result suggests that pleasant fragrances are associated with mild sweetness rather than intense sensory stimulation. The lactone components in fig scents provide a "milky" aroma that triggers maternal nursing memories, potentially inducing the brain to secrete oxytocin, which results in a sense of stable happiness. In fragrance design and marketing, fig notes are frequently blended with warm wood or mild fruits to cultivate a comfortable and joyful atmosphere. Schicber et al. (2025) noted that fruity scents consistently trigger feelings of pleasure and are widely utilized in Western cosmetics. However, a cross-cultural contrast emerges regarding spatial fragrances: for East Asian users, while fruity scents induce pleasure, they are not the primary consideration when selecting a fragrance for their environment. This

contrasts with Western consumers, who place a higher premium on high-pleasure, high-arousal fragrance profiles.

5.2 Analysis of High Pleasure and Low Arousal (LAP) Dimensions

Based on Russell's Pleasure-Arousal (P-A) matrix, the emotional states of Relaxation, Warmth, and Security are categorized within the High Pleasure and Low Arousal (LAP) area. This quadrant represents states of tranquility and emotional restoration, which are highly valued in East Asian "ideal affect" profiles.

5.2.1 Relaxation

The results identified "Relaxation" as the most influential Kansei construct (R = 0.524), indicating that emotional restoration constitutes the primary psychological value underlying fragrance attractiveness among Taiwanese users. Within the EGM hierarchy, "Forest Greenery Ambiance" served as the middle-level evaluation item, while "Moist Moss Scent" was identified as the dominant lower-level attribute. This hierarchical structure demonstrates how abstract emotional expectations are translated into concrete fragrance characteristics.

Natural scents have been shown to reduce stress (Taufer et al., 2025). In Taiwan, moss is culturally associated with silence and nature through Wabi-sabi aesthetics, explaining why Relaxation emerged as the strongest Kansei construct.

5.2.2 Warmth

Warmth (R = 0.381) was primarily supported by the middle-level evaluation item "Cozy Domestic Ambiance", with "Christmas Dessert Scent" identified as the dominant lower-level attribute. According to Reid et al. (2014), such nostalgic scents trigger the hippocampus within the limbic system, evoking positive memories associated with family and festivities. This psychological "Social Connectedness" translates into a perception of mental warmth, compensating for a sense of coldness or isolation in the physical environment.

While Christmas culture originates in the West, globalization has integrated these celebrations into Eastern life. Furthermore, while Westerners associate cinnamon and cloves with the holidays (Seo et al., 2009), these ingredients hold a parallel significance in Eastern culture as "warming" medicinal herbs (TCM) used in winter. Traditional Eastern desserts consumed during the Lunar New Year also feature sugary and spiced notes, triggering memories of "Reunion" (闔家團圓). Thus, "Warmth" in the East Asian context is a synthesis of globalized festive symbols and traditional culinary nostalgia.

5.2.3 Security

The EGM results reveal that the concrete features (lower level) associated with Security focus more on product descriptions than on specific olfactory notes. Most respondents believe that for a scent to provide security, it must faithfully represent product labeling.

Doris Schicker et al. pointed out that due to the interaction between olfactory and verbal processing regions in the brain, scents that align with their labels influence product perception more positively than labels alone. This semantic congruence—the relationship between the

linguistic descriptor and the scent itself—alters the consumer's psychological interpretation and evaluation (Iseki et al., 2021). In East Asian markets, reducing the gap between olfactory perception and the brand's narrative is crucial for enhancing "Security" and, consequently, increasing purchase intention.

Unlike the other affective constructs, Security is supported primarily by semantic consistency rather than by a specific fragrance note. This finding suggests that Kansei perception is influenced not only by sensory characteristics but also by the congruence between product communication and olfactory experience. Therefore, fragrance design should integrate scent development with packaging narratives and product descriptions to strengthen users' psychological trust.

6 CONCLUSION

This study demonstrates that integrating Miryoku Engineering with Quantification Theory Type I and the Pleasure–Arousal model provides a systematic framework for translating culturally specific fragrance perceptions into actionable design knowledge. While "Freshness" (HAP) enhances environmental invigoration, "Relaxation" (LAP) is the primary consumer preference, supported by "Wabi-sabi" aesthetics and nature-inspired scents. "Warmth" and "Security" are further shaped by cultural nostalgia and scent–label congruence.

Table 3 translates the hierarchical relationships identified through Miryoku Engineering into actionable fragrance design guidelines. Designers can use the identified lower-level fragrance attributes to support targeted Kansei values during product development. These findings provide a foundation for a systematic Kansei Design Framework linking subjective scent perception with product development. Future studies may extend the proposed framework to multisensory Kansei design by incorporating visual and tactile attributes.

Table 3. Design Guideline Matrix

Kansei Value	Evaluation Item	Fragrance Attribute	Design Recommendation
Freshness	Natural greenery	Moss	Botanical green accords
Pleasure	Sweet ambiance	Fig	Mild fruity fragrance
Relaxation	Forest greenery	Moist moss	Earthy green fragrance for bedrooms
Warmth	Cozy home	Christmas dessert	Vanilla + cinnamon notes
Security	Unpretentious natural ambiance	Scent profiles that faithfully reflect product labeling	Maintain scent–label consistency

REFERENCES

1. Chiu, T.-P. (2010). The study of emotion between scents and representational character of product design (Master's thesis, National Cheng Kung University). Institutional Repository.
2. Kiecolt-Glaser, J. K., Graham, J. E., Malarkey, W. B., Porter, K., Lemeshow, S., & Glaser, R. (2008). Olfactory influences on mood and autonomic, endocrine, and immune function.

Psychoneuroendocrinology, 33(3), 328–339. https://doi.org/10.1016/j.psyneuen.2007.11.015

3. Müller, J. (1992). The H&R book of perfume: Understanding fragrance—Origins, history, development; Guide to fragrance ingredients (1st ed., pp. 66–74). Glöss Verlag.
4. Nagamachi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. Applied Ergonomics, 33(3), 289–294. https://doi.org/10.1016/S0003-6870(02)00019-4
5. Reid, C. A., Green, J. D., Wildschut, T., Sedikides, C., & Parsons, E. (2015). Scent-evoked nostalgia. Memory, 23(2), 157–166. https://doi.org/10.1080/09658211.2013.876048
6. Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714
7. Sanui, J. (1996). Visualization of users' requirements: Introduction of the evaluation grid method. In Proceedings of the 3rd Design & Decision Support Systems in Architecture and Urban Planning Conference (pp. 365–374).
8. Schicker, D., Xie, Y., Springer, A., Karacan, B., & Freiherr, J. (2025). Smell the label: Odors influence label perception and their neural processing. Journal of Neuroscience, 45(14), e1159242024. https://doi.org/10.1523/JNEUROSCI.1159-24.202
9. Seo, H.-S., Buschhüter, D., & Hummel, T. (2009). Odor attributes change in relation to the time of the year: Cinnamon odor is more familiar and pleasant during the Christmas season than summertime. Appetite, 53(2), 222–225. https://doi.org/10.1016/j.appet.2009.06.011
10. Taufer, F., Cortese, L., Cattaneo, A., et al. (2025). Psychological and physiological responses to smells from nature: Potential health benefits for urban dwellers. npj Urban Sustainability, 5(1), Article 27. https://doi.org/10.1038/s42949-025-00274-0
11. Thiboud, M. (1991). Empirical classification of odors. In P. M. Müller & D. Lamparsky (Eds.), Perfumes: Art, science, and technology (pp. 253–286). Elsevier.
12. Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. Perspectives on Psychological Science, 2(3), 242–259. https://doi.org/10.1111/j.1745-6916.2007.00043.x
13. Ujigawa, M. (1997). The practice of Miryoku engineering. Kaibundo.
14. Zarzo, M. (2013). What is a fresh scent in perfumery? Perceptual freshness is correlated with substantivity. Sensors, 13(1), 463–483. https://doi.org/10.3390/s130100463

DESIGNING FOR ENRICHING LEISURE EXPERIENCE WITH A CRAFTED TEA SET OF EMBODIED CULTURAL STORIES

Hsuan-Hsuan Su [a], Yen-Chun Lin[a] , Ya-liang Chuang[b*1] , Fu-Chi, Shin[c], Yun-Wen Li[d]

[a] *National Tsing Hua University, Hsinchu, 2hsuanceramics@gmail.com*
[a] *National Tsing Hua University, Hsinchu, linyanjun168@gmail.com*
[b] *National Tsing Hua University, Hsinchu, yaliang@mx.nthu.edu.tw*
[c] *National Tsing Hua University, Hsinchu, fuchi@mx.nthu.edu.tw*
[d] *National Tsing Hua University, Hsinchu, leeyw@mx.nthu.edu.tw*

ABSTRACT

In today's fast-paced lifestyle, people increasingly prioritize the quality of leisure time with family and friends for emotional restoration. Research indicates that low-effort activities during non-work hours help reduce burnout; however, such activities often lack sensory depth and interpersonal interaction. This study focuses on "weekend gatherings" moments and proposes a tea set design that utilizes craftsmanship and cultural elements to enrich users' emotional experiences and social connections. Our design process combined both quantitative and qualitative methods. Firstly, we applied Kansei Engineering Type I approach to investigate users' perceptions towards 15 existing tea set products. 33 participants completed a survey of 10 semantic differential scales. Factor analysis identified three key Kansei factors: "Ritual," "Innovation/Interest," and "Decoration." Informed by those insights, we integrated cultural elements into the ritual of tea time design and create a ceramic tea set that offers rich sensorial experiences. In the post-test, we invited 5 participants to try out our prototype in a social gathering setup to understand their user experiences through observation and semi-structured interviews. The results show that: (1) the puzzle game of the ritual design successfully triggered surprise and curiosity; (2) the integration of multisensory elements throughout the tea-making process enhanced participants' engagement, relaxation, and emotional restoration; (3) the combination of craftsmanship, cultural elements, and ritualized interaction strengthened participants' emotional connection and social engagement during weekend tea gatherings.

Keywords: *Kansei Engineering, Multi-sensory Experience, Craftmanship, Emotional Restoration, Tea Set Design.*

[1] Corresponding author: Ya-Liang Chuang. E-mail:yaliang@mx.nthu.edu.tw

1 INTRODUCTION

1.1 Research Background and Motivation

Under the high-pressure lifestyle of the modern era, weekends serve as a critical "recovery" period. Engaging in voluntary leisure activities helps restore depleted psychological and physical resources (Jeong et al., 2020). Weekend gatherings, categorized as "joint leisure," provide a space for social support and emotional sharing (Rimé, 2007). In such contexts, craft objects act as essential mediators that shape the atmosphere and collective memory.

However, in traditional craft practice, the evaluation is often skewed toward technical precision and material mastery, frequently neglecting the "Kansei intention" of the creator and the actual "perceived experience" of the user. This gap results in objects that are technically refined but emotionally disconnected. Therefore, this study introduces Kansei Engineering Type I to systematically map physical attributes to psychological perceptions, ensuring that craft materials—through touch, weight, and texture—effectively drive emotional restoration.

1.2 Research Objectives

The primary objective of this study is to establish a systematic development model for craft design that bridges the gap between creator intention and user perception. Specifically:

1. To identify key Kansei factors for tea sets in social contexts using the SD method.

2. To translate Kansei factors into concrete formal, sensory, and interaction design parameters for the development of the tea set.

3. To validate the effectiveness of the proposed tea set design in enhancing relaxation and emotional connection during weekend tea gatherings.

2 LITERATURE REVIEW

2.1 Multi-sensory Experience and Holistic Perception

In contemporary design research, multisensory experience has emerged as a pivotal theoretical lens for deciphering the emotional nexus between users and objects. Unlike analytical frameworks centered on isolated sensory modalities, multisensory design posits that human perception is an integrated experiential process; the simultaneous engagement of multiple senses shapes the holistic cognition and emotional response toward an object (Yin, Zhu & Yuan, 2024).

According to the Kansei Potential Thinking Method (Murata, 2014), humans do not process external stimuli through discrete sensory channels; instead, they form a "holistic perception" through sensory integration. This concept suggests that information from visual, tactile, and olfactory channels converges at the cognitive level to construct a comprehensive understanding and Kansei judgment of the object. Within the context of craft and product design, the formal language, material properties, and interaction sequences serve as critical design parameters that modulate this holistic perception.

Extant research underscores that Kansei value is a decisive factor in stimulating user interest, cultivating preferences, and driving decision-making, thereby facilitating emotional resonance among the designer, the object, and the user (Norman, 2004).

From the perspective of User Experience (UX), vision typically functions as the primary sensory receptor, responsible for the rapid appraisal of form, material, proportion, and aesthetic composition. However, compared to the immediate impressions provided by vision, tactile perception is regarded as the most fundamental pathway to perceiving the essence of materials. Through feedback from the skin and musculature, users discern physical attributes such as weight, temperature, dimensions, and texture (smoothness vs. roughness), thereby establishing a profound understanding of material and craft quality (Tiest, 2010). Consequently, Kansei value is not merely generated through observation but must be accumulated and internalized through physical engagement, manipulation, and iterative use—a process rooted in embodied experience. In this context, multisensory design not only enriches the experiential depth but also transforms mundane daily rituals—such as tea tasting—into collective experiences characterized by emotional gravity and memorability.

2.2 Kansei Craft and Emotional Value

The term "craft" can be traced back to the Arts and Crafts Movement. In the era of mass production following the Industrial Revolution, most products relied on mechanized manufacturing, causing objects to lose the delicate and warm texture inherent in handmade work—often critiqued as design lacking "soul" or emotional resonance. Consequently, craft design integrates manual practice into the design process, endowing objects with formal warmth and diversity. In his study on "craft consumption," Bell (2021) posits that consumers pursue craft objects not merely for the manual process or the identity of the maker, but because the consumption process involves a synthesis of craft judgment, emotional investment, passion, and the intrinsic allure of the object. These elements constitute a conscious practice through which consumers express self-identity and create value. Thus, through the praxis of materials, craft ensures that a work functions not only as a utilitarian tool but also as a vital medium for shaping form and meaning while responding to life and contemporary experiences.

Among the five senses, vision is often regarded as the primary intuition for appraising an object's aesthetics, whereas tactile perception is considered the most fundamental pathway to experiencing the material itself. Through cutaneous feedback, individuals can discern characteristics such as size, weight, and surface texture (smoothness vs. roughness), thereby absorbing extensive information regarding the object (Tiest, 2010). Similarly, Yanagi (2002) noted in his theory of folk art (Mingei) that "to understand the beauty of an object, one must first touch its materials and textures with their own hands" (p. 182). Therefore, it is only through the physical proximity to materials—accumulating perceptual experience through contact and use—that the distance between human and object is bridged, allowing the Kansei value embedded in craft to truly manifest.

3 METHODOLOGY

This research aims to address the potential perceptual discrepancy between the Kansei intentions conveyed by designers and the actual perceived experiences of users within the practice of craft design. Using the design of a tea set for 'weekend gatherings' as a primary case study, this paper employs a systematic methodology to establish a mapping relationship between product formal attributes and user Kansei values. The findings serve as a robust theoretical framework for informed design decision-making in subsequent development processes.

The first phase aimed to establish the Kansei positioning of the target product and apply Kansei Engineering to translate users' subjective emotional responses into analyzable design parameters. The research procedure consisted of competitive product analysis, Kansei word construction, a Semantic Differential (SD) survey, and Kansei space construction.

First, approximately 30 tea sets, tableware products, and craft artworks were collected from the market. Based on differences in product style, usage context, and formal characteristics, together with the six dimensions of Kansei value proposed by Murata (2020), a qualitative analysis was conducted to select 15 representative samples for further evaluation. These samples were chosen to represent the major design languages and emotional characteristics found in the current market.

Next, a Semantic Differential (SD) scale was developed using ten pairs of Kansei adjective scales to evaluate users' perceptions of the products in terms of visual impression, usage context, and emotional atmosphere. Representative adjective pairs included Soft–Vivid, Practical–Decorative, Interesting–Serious, and Ritualistic–Everyday, among others

An online questionnaire survey was then conducted, yielding 33 valid responses. Participants evaluated each representative product using a seven-point Likert scale based on the ten Semantic Differential adjective pairs. The collected data were analyzed using factor analysis to extract the principal Kansei dimensions, construct the Kansei space, and perform product positioning analysis. As shown in Table 1 and Table 2.

The results identified three primary Kansei factors underlying users' perceptions of the representative products: Decoration, Innovation/Modernity, and Ritual. These factors subsequently served as the foundation for translating users' emotional expectations into the formal, visual, and interaction design parameters of the proposed tea set.

Table 1. Total Variance Explained

Total Variance Explained						
Initial Eigenvalues				Extraction Sums of Squared Loadings		
Component	Total	Variance%	Cumulative %	Total	Variance %	Cumulative %
1	5.270	52.701	52.701	5.270	52.701	52.701
2	2.289	22.893	75.594	2.289	22.893	75.594

3	1.311	13.110	88.704	1.311	1.311	88.704
4	.513	5.129	93.833			
5	.317	3.169	97.002			
6	.129	1.286	98.287			
7	.085	.846	99.134			
8	.049	.492	99.625			
9	.036	.360	99.986			
10	.001	.014	100.000			
Extraction Method: Principal Component Analysis						

Table 2. Component Matrix

Component Matrix[a]			
Component			
	1	**2**	**3**
Refined – Rustic	-.877	-.299	-.007
Minimal – Elaborate	.814	.470	-.064
Simple – Ornate	.811	.503	.129
Practical – Decorative	.745	.002	-.552
Natural – Artificial	.725	-.071	.539
Soft – Bright	.702	.066	.625
Sense of ritual-Everday	-.696	-.378	.508
Traditional – Modern	.685	-.674	.021
Novel – Vintage	-.563	.800	.025
Fun – Serious	-.578	.692	.212
Extraction Method: Principal Component Analysis a.3 components extracted.			

3.1 Phase I: Quantitative Analysis (Pre-test)

- Sample Selection: 15 representative tea set designs were selected.
- SD Scale: 10 pairs of bipolar Kansei words (e.g., Ritualistic–Everyday, Innovative–Traditional).
- Subjects: 33 participants (N=33) evaluated samples using a 7-point Likert scale.
- Data Analysis: Factor analysis was performed to construct a Kansei imagery space.

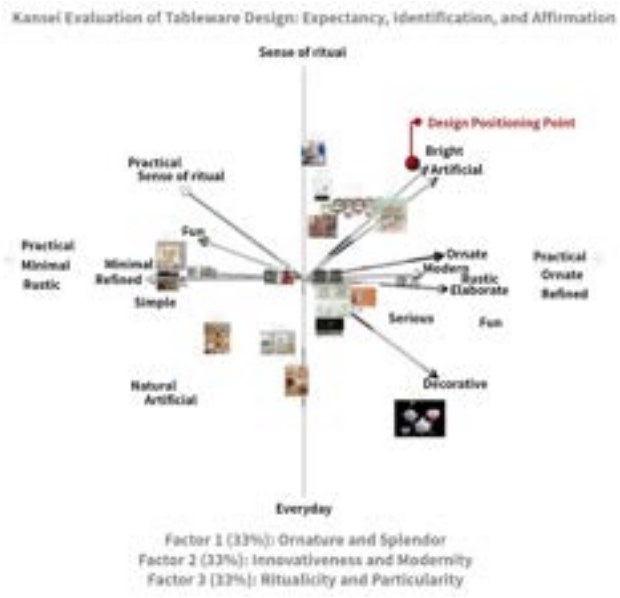


Figure 1. Design Position within the Affective Space

3.2 P Phase II: Design Implementation and Qualitative Evaluation (Post-test)

Based on the factor analysis, a "Floral Language" tea set was developed focusing on high "Ritual" and "Innovation."

- User Testing: 5 subjects participated in a simulated "weekend gathering."
- Evaluation Tools: Behavioral observation and semi-structured interviews.
- Analysis: Affinity Diagram (KJ Method) was used to categorize user feedback into Visual/Psychological, Sensory, and Ergonomic dimensions.

Based on the results of the Semantic Differential evaluation, three major Kansei dimensions were identified: Ritual, Innovation/Interest, and Decoration. These Kansei factors guided the translation of users' emotional expectations into specific design decisions. Ritual was reflected through the sequential tea-making process and concealed interactive mechanisms; Innovation/Interest was represented by hidden decorative surprises beneath the cup and saucer; Decoration was expressed through hand-painted illustrations and visual details. The optional wood chips were introduced only to complement the overall experiential atmosphere and were not directly derived from any individual Kansei factor.

4 RESULTS AND DISCUSSION

4.1 User Experience Observation

To further understand users' experiences with the ceramic tea set in a social gathering setup, this study employed scenario simulation combined with qualitative methods. Five participants were invited to try out the prototype within a pre-defined "weekend gathering" scenario. Their experiences were examined through non-participatory observation and semi-structured interviews, with particular attention given to their sensory reactions, emotional responses, and the fluidity of the operational process.

During the observation process, the researchers focused on participants' initial responses when interacting with the tea set, as well as how they operated and perceived the overall tea-time ritual. The results showed that when participants lifted the teacups and discovered the hidden patterns and wood chips beneath the saucers, most of them immediately showed surprise and curiosity. This indicates that the puzzle game of the ritual design successfully triggered the "Innovation/Interest" Kansei factor identified in the previous quantitative analysis.

Furthermore, during the tea-tasting process, several participants actively smelled the wood chips and showed relaxed body language, such as deep breathing and slower movements. This suggests that the integration of olfactory elements, especially the aroma from wood chips, could enhance relaxation and support emotional restoration. The findings also indicate that rich sensorial experiences, including visual, tactile, and olfactory stimuli, can deepen users' emotional experiences beyond visual appreciation alone.

At the same time, when participants operated the ceramic heating base and candlelight apparatus, their actions became slower, more careful, and more focused. This suggests that the combination of craftsmanship, heat, and cultural elements helped transform tea drinking from a simple daily activity into a meaningful tea-time ritual. In this way, the ceramic tea set not only enriched sensory depth but also encouraged interpersonal interaction and social connection within the weekend gathering context.

4.2 Qualitative Synthesis of Semi-Structured Interviews via Affinity Diagram

Following the interaction experience, semi-structured interviews were conducted to understand participants' feelings and feedback. The interview data were analyzed using the Affinity Diagram, also known as the KJ Method, to organize the feedback into different levels and identify the main parts of users' Kansei experiences. The results categorized user feedback into the following two main dimensions:

1. Visual and Psychological Dimension: Imagery Translation and Exploratory Surprise Participants generally perceived that the concealed design elements and meticulous detailing successfully translated the designer's intended imagery, significantly enhancing the product's sense of surprise and explorability. Concurrently, the hand-painted motifs of lavender, chrysanthemum, and osmanthus allowed participants to perceive the intricacy of the craftsmanship, thereby elevating their appraisal of the product's perceived refinement.

2. Sensory Dimension: Therapeutic Experience through Cross-Media Integration The therapeutic experience derived from cross-media integration emerged as a focal point of consensus among participants. It was unanimously noted that the fragrance emitted by the heated wood chips—and its associated symbolic meaning—effectively mitigated stress and cultivated a relaxed, intimate atmosphere within the social setting. Furthermore, the moderate thermal conductivity of the ceramic vessels provided supportive tactile feedback, imbuing the overall experience with a sense of warmth and comfort.

Summary of Empirical Findings By synthesizing the results of behavioral observations and the Affinity Diagram analysis, this study further developed a hierarchical tree structure of the user's

Kansei experience. This construct serves as a critical theoretical reference for the subsequent optimization of craft design and the strategic calibration of multi-sensory design interventions.

5 RESULTS AND DISCUSSION

5.1 Summary and Future Recommendations

The results confirmed that the three Kansei factors identified through factor analysis—Ritual, Innovation/Interest, and Decoration—could be further translated into corresponding design elements of the tea set. Specifically, the heating base and candlelight apparatus jointly created a sense of ritual during the tea-serving experience. The concealed structure containing wood chips and the thoughtful concept of using the tea set as a gift enhanced users' perceptions of innovation and interest. Meanwhile, the hand-painted illustrations further strengthened the tea set's decorative qualities.

The findings also demonstrated that Kansei Engineering Type I can assist designers in translating users' emotional semantics into concrete form, structural, and interaction design features, thereby reducing the gap between the designer's intended emotional expression and the users' actual perceptual experience. Therefore, Kansei Engineering serves not only as a systematic method for translating Kansei semantics into design elements but also as a structured framework for supporting the development of craft design.

Furthermore, the integration of multisensory elements enriched the sensory and emotional experience of the ceramic tea set. The subtle aroma released from the wood chips contributed primarily to relaxation and emotional restoration, while the candlelight, warmth, and heating apparatus collectively created a contextual atmosphere that reinforced the sense of ritual during the tea gathering. These findings suggest that, compared with designs focusing solely on visual aesthetics, a multisensory design approach can more effectively enhance users' engagement and emotional connection throughout their interaction with the tea set.

5.2 Research Scope and Limitations

1. Contextual Scope: This study focuses exclusively on small-scale "weekend gathering" scenarios. The discussion centers on the impact of holistic tea set forms on atmosphere and emotional perception; it does not encompass professional tea ceremonies, commercial hospitality spaces, or long-term daily usage.
2. Temporal Constraints of Simulation: User experience data were collected during short-term simulated scenarios. As a "single-session gathering," the study does not account for habit formation, material wear, or emotional attachment accumulated over extended periods. Therefore, the appraisals primarily reflect immediate, first-contact perceptions.
3. Material Experimental Boundaries: The material selection and craftsmanship were limited to ceramic media, excluding a wider range of glaze formulations or cross-material combinations. Consequently, the discussion on the "Kansei value of craft materials" represents a preliminary exploration within a specific material set and may not fully represent the broader craft domain.

ACKNOWLEDGMENTS

We would like to thank our advisors and course instructors in our department for their valuable guidance and suggestions on the research direction and manuscript preparation. We also thank our research partners for their support in research discussions and data organization. Finally, we thank all participants in this study, and KEER for providing the opportunity for academic exchange and presentation.

REFERENCES

1. Adamson, G. (2013). The invention of craft. Bloomsbury Academic.

2. Bergmann Tiest, W. M. (2010). Tactual perception of material properties. Vision Research, 50(24), 2775–2782. https://doi.org/10.1016/j.visres.2010.10.005

3. Chuang, M. C., & Lin, Y. T. (2013). Color schemes on different media and their imagery transmission [Master's thesis, National Chiao Tung University]. Google Books.

4. Greenhalgh, P. (1997). The persistence of craft: The applied arts today. Rutgers University Press.

5. Huang, W. H. (2011). On the "Common Sense" in modern Japanese philosophy. Bulletin of the College of Liberal Arts, National Taiwan University, (5), 217–241.

6. Jeong, J.-G., Kang, S.-W., & Choi, S. B. (2020). Employees' weekend activities and psychological well-being via job stress: A moderated mediation role of recovery experience. International Journal of Environmental Research and Public Health, 17(5), 1642. https://doi.org/10.3390/ijerph17051642

7. Liu, C. Y., & Tang, L. Y. (2009). Integrating grey relational analysis and Kansei engineering into product design and development. Journal of Quality, 16(3), 165–177.

8. Murata, C. (2020). Kansei potential thinking: Business innovation from zero. China Productivity Center. (Original work published 2014).

9. Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books. https://doi.org/10.5860/CHOICE.41-6846

10. Rimé, B. (2007). Interpersonal emotion regulation. In J. J. Gross (Ed.), Handbook of emotion regulation (pp. 466–485). The Guilford Press.

11. Yanagi, S. (2002). The theory of folk art (Mingei). Jiangxi Art Publishing House.

DESIGNING FOR RITUALS:
CALIBRATING SLOW INTERACTION IN SOCIAL CONTEXTS WITH AN INTERACTIVE COASTER

Ya-Ching Chang[1], Shang-Yuan Chang[2], and Yaliang Chuang[3]

[1] *National Tsing Hua University, Hsinchu, Taiwan, yaching.chang@gapp.nthu.edu.tw*

[2] *National Tsing Hua University, Hsinchu, Taiwan, nn113593519@gapp.nthu.edu.tw*

[3] *National Tsing Hua University, Hsinchu, Taiwan, yaliang@mx.nthu.edu.tw*

ABSTRACT

Many interactive systems are optimized for efficiency, yet everyday rituals, such as drinking, often benefit from friction and deliberate pacing. Slow Design offers an alternative lens for this inquiry. By introducing behavioral friction—resistance that demands conscious engagement—it naturally triggers the appraisal process for reflection, rather than simply switching a function on or off. This study investigates how such friction can be embedded into everyday objects and how people will experience it in different social contexts. We combined the research-through-design approach and Kansei Engineering methods in designing an interactive coaster across three phases. First, we collected a diverse set of coaster samples and selected 15 representative samples to evaluate users' emotional perceptions using a Semantic Differential survey (N=32). The Exploratory Factor Analysis yielded three key factors that accounted for 88% of the variance. They are *emotional expression and style*, *technology and manufacturing orientation*, and *value and durability*. Second, guided by these findings and Slow Design principles, we developed an NFC-enabled interactive coaster offering multi-sensory feedback through ambient LED lighting and an AI-curated application for mood logging and shared history. Third, we evaluated the prototype in two situated contexts: a commercial bar (N=5) and a domestic setting (N=3). Participants in both contexts valued the coaster's tactile materiality and ambient light; in solitary use, high-friction mood logging turned waiting into a "healing" ritual, and ambient cues opened conversations. Explicit app interaction, however, interrupted group conversation, revealing a core tension: friction that affords reflection in solitude becomes disruption in company. We contribute three design provocations—Gentle Revelation from the reach of peripheral cues, Reflective Pacing from the divergent reception of enforced delay, and Contextual Interaction Friction from the tension itself—reframing friction as a deliberate affordance that must be dynamically calibrated to social context, not universally prescribed.

Keywords: *Slow Design, Slow Technology, Tableware, Interactive Design*

1 INTRODUCTION

In contemporary ubiquitous computing, product design prioritizes efficiency and friction reduction. While these priorities have yielded highly optimized "smart" systems, they often inadvertently strip daily routines of the friction necessary for mindfulness, reducing meaningful human–object interactions to instrumental tasks (Hassenzahl, 2010; Hassenzahl et al., 2010; Norman, 2004). This motivates the deliberate reintroduction of meaningful friction and temporal pacing into routine interactions to support contemplation and self-awareness.

Research in slow design (Fuad-Luke, 2002) and slow technology (Hallnäs & Redström, 2001) has advocated for interaction qualities that introduce pauses, anticipation, and material engagement to support mindfulness and emotional awareness. Strauss and Fuad-Luke (2008) articulate six core principles of Slow Design, which serve as the theoretical and practical framework: (i) Reveal: Uncovering often-overlooked processes, resources, and impacts within daily life, (ii) Expand: Extending the implications of design into broader social and cultural contexts, (iii) Reflect: Facilitating user introspection regarding their own behaviors and values, (iv) Engage: Encouraging active engagement rather than passive consumption, (v) Participate: Emphasizing co-creation between designers and users, and (vi) Evolve: Allowing design to mature and develop over time, rather than being finalized at the point of manufacture.

Through Kansei Engineering (KE) analysis, designers can identify and select material parameters that accumulate emotional value over time, thereby translating abstract slow-design principles into concrete, executable design strategies (Schütte, 2005). The distinction between process and outcome is central to our theoretical framing. Unlike Reflective Design, which primarily targets cognitive introspection as its end goal, Slow Design operates at the behavioral level first: deliberate material resistance and interaction friction break automaticity, triggering the appraisal process through which reflection becomes possible (Odom et al., 2022; Strauss & Fuad-Luke, 2008). This grounding guided our semantic differential (SD) survey (Osgood, 1957) and the EFA-guided design space toward a coaster with the qualities of emotional richness, high interactivity, and humor—the three underserved dimensions in which slow design could be embedded into a daily artifact.

In design practice (Cox et al., 2016; Grosse-Hering et al., 2013; Marsh, 2016), however, much of the existing literature assumes that slowness is inherently positive, overlooking how "slow interactions" are perceived across varied social contexts. To address this gap, based on the design-identified space, we applied a Research through Design (RtD) approach (Zimmerman et al., 2010) to develop and evaluate a slow-designed interactive coaster in situated social contexts.

Taken together, the central research question guiding this study is: (1) how can behavioral friction be deliberately embedded into a familiar everyday object to support ritual and reflection, and (2) how is its value shaped by social context? We address this through two sub-questions: (RQ1) How can Kansei Engineering and Slow Design principles translate abstract “slowness” into a concrete interactive artifact? (RQ2) Does the friction that fosters mindful reflection in solitary use disrupt the flow of interaction in group settings?

The findings yield three design provocations (Odom, 2024). Rather than treating friction as an obstacle, these provocations—Gentle Revelation, Reflective Pacing, and Contextual Interaction Friction—reframe it as a deliberate affordance whose demands are calibrated to social context, moving slow design from the philosophical toward the actionable.

2 METHODS

To operationalize slow design principles, this study adopted a Research through Design (RtD) (Zimmerman et al., 2010) approach spanning three phases: (i) quantitative Kansei Engineering, (ii) prototype development, and (iii) qualitative design evaluation.

2.1 Phase I: Quantitative Kansei Engineering

2.1.1 Sample Selection and Instrument Development

To understand drink-related rituals in everyday life, we conducted field observations at various venues (e.g., bars, coffee shops) and domestic settings. Observing how people interacted with tableware informed our design brainstorming and led to the curation of 15 distinct coaster samples (see Figure 1). To evaluate user perceptions and interaction experiences with these artifacts, we developed a semantic differential (SD) survey (Osgood, 1957) with 10 selected bipolar adjective pairs.



Figure 1: The fifteen diverse coaster design samples collected and used in our survey to discover unmet user needs.

The initial bipolar adjective pairs were selected from the product aesthetics vocabulary proposed by Johnson et al. (2003) and mapped through the lens of slow design theory (Strauss & Fuad-Luke, 2008). The adjective pairs were predominantly collected from the 'Perception-Symbolic' category in Johnson et al. (2003)'s work. This concentration is a deliberate methodological choice, as slow design is fundamentally a meta-cognitive approach. It aims to uncover hidden processes (Reveal), trigger user introspection (Reflect), extend cultural implications (Expand), demand active cognitive involvement (Engage), emphasize designer-user co-creation (Participate), and allow the object's meaning to mature over time (Evolve), rather than merely manipulating physical 'Aesthetics-Sensory' attributes.

Furthermore, a pilot test with 5 participants indicated that strict adherence to the original academic terminology (e.g., Ornamental vs. Functional) caused ambiguity. Consequently, the survey was adapted to utilize more accessible vocabulary (e.g., Emotional vs. Functional). The finalized 10-item scale operationalizes five of the six slow design principles (see Appendix)—

excluding Participate, which is defined as a co-creation process during the development phase rather than as perception during the use phase. Finally, 32 participants rated the 15 coaster samples using this adapted 10-item, seven-point Likert-type scale.

2.1.2 Procedure and Data Analysis

After conducting the semantic differential rating, we applied Exploratory Factor Analysis (EFA) to condense the semantic ratings into a set of underlying Kansei factors. The EFA established a three-factor structure: Emotional Expression and Style (Factor 1), Technology and Manufacturing Orientation (Factor 2), and Value and Durability (Factor 3) (see Table 1), accounting for 87.94% of the variance. We then plotted the 15 coaster samples across the three-factor space using a three-dimensional rotation map (see Figure 2), which visualizes each sample's position relative to all three Kansei factors simultaneously. By examining which region of this space was underserved—specifically, artifacts that score high on emotional richness, high interactivity, and humor remain rare among existing products—we derived the unmet design opportunity that guided prototype development.

Table 1. Rotated Component Matrix of exploratory factor analysis (EFA)

Bipolar Adjective Pairs	Solution 1	Solution 2	Solution 3
Low-tech vs. High-tech	0.923	-0.046	0.313
Handmade vs. Manufactured	0.908	0.041	-0.261
Low Interactive vs. High Interactive	0.908	-0.199	0.02
Classic vs. Trendy	0.795	-0.475	-0.108
Emotional vs. Functional	0.269	0.901	-0.076
Humorous vs. Humorless	-0.324	0.792	0.318
Monotonous vs. Rich	0.324	-0.739	0.292
Novel vs. Ordinary	-0.625	0.7	-0.248
Durable vs. Disposable	0.258	-0.041	-0.893
Cheap vs. Expensive	0.489	-0.225	0.81

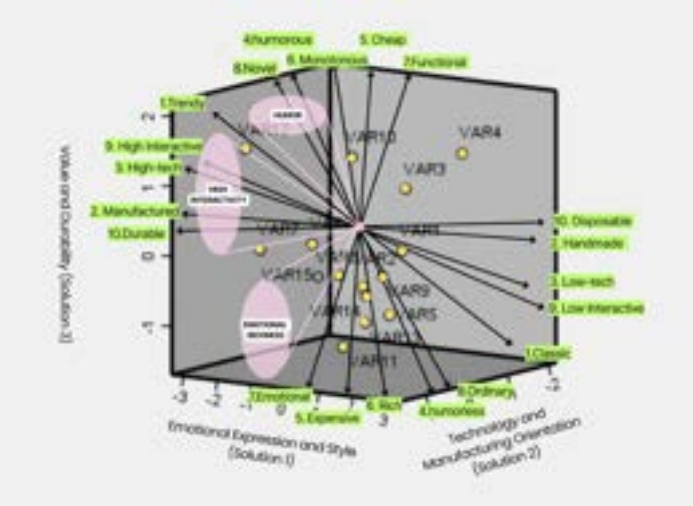


Figure 2. Mapping the EFA result in three-dimensional space reveals the unmet design opportunity.

2.2 Phase II: Prototype Development and Slow Design Integration

Guided by the findings in Phase I and Slow Design principles (Strauss & Fuad-Luke, 2008), we developed the "Slow Coaster," an NFC-enabled artifact designed to introduce meaningful friction into the drinking ritual.

2.2.1 The Physical Artifact: Visceral Friction

The coaster is handcrafted from irregular glass, a material choice driven by the slow design principles (Strauss & Fuad-Luke, 2008) of Reveal and Reflect. Unlike the smooth uniformity of industrial plastic, the glass's uneven texture and weight distribution introduce tactile "visceral friction," forcing the user to handle the object with conscious care rather than mindless automaticity (Pallasmaa, 2024). Crucially, the LED light embedded in the coaster provides ambient feedback and high interactivity at the visceral level (Chuang, 2020; Norman, 2007). By compelling the user to handle the object with focused sensory engagement—consisting of tactile and subtle visual cues—the design shifts the interaction from a mindless placement to a conscious, mindful ritual (Karana, Barati, et al., 2015; Karana, Pedgley, et al., 2015).

2.2.2 The Digital Interface: Reflective Pacing

The NFC tags embedded under the coaster link the physical artifact to a web-based application designed for Reflection (Strauss & Fuad-Luke, 2008). The system challenges the "instant access" paradigm through the following mechanisms:



Figure 3. The main interactions of using the interactive coaster prototype developed in this study: (A) Launch the App by tapping a smartphone on the coaster. (B) The App launches at a deliberate pace to slow down a user's mind. (C) A user reflects on their current mental state and logs their mood. (D) The system sets a corresponding atmosphere with matched light color and music. (E) A user can also revisit their past reflections and listening history. He/she can also browse other people's publicly shared data.

High-Friction Entry: Users must physically scan the coaster to launch the app (see Figure 3A), breaking the stimulus-response loop of standard notifications.

Temporal Buffering: The app interface purposefully loads at a deliberate pace (see Figure 3B), transforming the wait time into a moment of anticipation and mental transition

Emotional Mirroring: Users log their current mood (see Figure 3C), which the system translates into matched ambient light colors on the coaster and an AI-curated playlist (see Figure 3D), externalizing internal states using the environment to mirror the user's emotions (Boehner et al., 2007; Norman, 2007).

AI and Social Connection for Humor: The app integrates a witty music recommendation engine based on the user's mood and written notes, and provides a 'digital footprint' (music/mood history) left by previous users of the same coaster (Figure 3E), fostering long-term emotional durability in a shared context (Chapman, 2012; Hassenzahl, 2010; Hassenzahl et al., 2010). In this study, multiple Spotify playlists were predefined to simulate AI functionality, each containing more than 10 songs that express a particular mood.

2.3 Phase III: Qualitative Design Evaluation

To evaluate how social context influences the experience of "slowness," we conducted a situated test in two contrasting environments: a lively commercial bar (N=5, one group setting and three solitary settings) and a quiet domestic living room (N=3, one group setting and one solitary setting) (see Table 2). The bar setting was deliberately chosen to introduce environmental complexity—noise, ambient lighting, and ongoing social dynamics—as a stress test for the design's contextual adaptability. Within this setting, professional staff (P1, bartender; P2, waitress; P5, bar owner) were intentionally recruited alongside regular customers to capture expert perspectives on how such a design might function in a real commercial deployment, including its impact on service flow and customer experience at scale. Consistent with the exploratory nature of Research through Design, this small purposive sample (N=8) prioritizes contextual richness over statistical generalizability.

Table 2. Participant overview for the Phase III design evaluation

Participant	Setting	Social Context	Role
P1	Bar	Solitary	Bartender
P2	Bar	Solitary	Waitress
P3	Bar	Group	Customer
P4	Bar	Group	Customer
P5	Bar	Solitary	Bar Owner
P6	Domestic	Group	Participant
P7	Domestic	Group	Participant
P8	Domestic	Solitary	Participant

Procedure: Participants were invited to use the coaster during a standard drinking session involving either tea, coffee, or alcohol. We employed a Wizard-of-Oz protocol to simulate the system's "emotional mirroring" light features, allowing researchers to trigger light changes in

response to user inputs manually, ensuring consistent timing. Participants were instructed to pour or order a drink, scan the coaster to launch the web app, log their current mood, and interact with the artifact while consuming their beverage.

Data Collection: We gathered qualitative data through direct observation during the 20-minute sessions and semi-structured post-study interviews. Our analysis focused on capturing micro-interactions—specifically, how users negotiated high-friction features (scanning, loading delays) amid the distractions of a bar versus the quiet of a home.

3 FINDINGS

The interview data were first transcribed, and two researchers independently coded the transcriptions. We then used thematic analysis (Braun & Clarke, 2006) to cluster the codes and identify the most important themes. Participants recruited in the bar context were coded as P1-P5, and those in the domestic context as P6-P8. In the bar setting, P1 (bartender), P2 (waitress), and P5 (bar owner) offered professional perspectives on the design. Participants P3 and P4 and P6 and P7 tested the design in group settings. During our analysis, we focused on emotional resonance, ritual formation, and the influence of social context on the acceptance of "slow" interaction features.

Our findings highlight a clear contrast in how "slowness" was perceived across social settings. While all participants in both contexts praised the physical (glass) and ambient (light) features, the explicit interactions (app/mood-logging) revealed a sharp divide between solitary introspection and social flow. The three primary themes are presented below.

3.1 Solitary Resonance: The "Healing" Ritual

In moments of solitude or quiet reflection, the system's high-friction design successfully fostered mindfulness. Participants described the experience as a "healing" process, with P2, P3, P6, and P8 noting that the pacing helped them enter a state of *"mental emptiness"* and *"relaxation."* Specifically, we observed the following two reflective behaviors.

Introspection: P1, P2, and P5 highlighted that the mood-logging feature facilitated an *"awareness of the present self."* P5 (Bar Owner) specifically noted that for solitary customers, the slow interaction transformed the idle time of waiting for a drink into a *"ritualistic experience,"* creating a psychological space for self-dialogue.

Reframing Friction: Although P6 and P7 initially found the NFC scanning and slow app loading unfamiliar and annoying, they ultimately reported that this friction forced them to *"be in the present."* During the testing, they were intrigued by the combination of tangible and digital interaction the design offered, realizing the purpose of slow design in bringing them to the present.

3.2 Social Dissonance: Friction as Disruption

In contrast, the same "slow" features often failed in group settings, where participants prioritized conversation over device interaction. This was the most consistently substantiated

finding in our evaluation: six of eight participants (P1, P2, P3, P4, P5, and P6) reported that slow interactions were undesirable in social contexts because they interrupted the natural flow of interaction among people. Essentially, it was due to the interruption of social flow, in which a user needs to focus on what is shown on their phone rather than on social interactions. P6 provided the most direct critique of this tension: *"When friends are there, I don't want to spend time selecting my mood—I want to interact with them."* This suggests that while the artifact (coaster) belongs on the table, the interaction (looking at a phone) signals social disengagement.

In addition, the professional participants (P1, P2, and P5) identified scalability issues in the two distinct contexts. P5 (Bar Owner) observed that while the experience creates an intimate atmosphere for couples or solo drinkers, it is *"inapplicable to large-scale group gatherings"* where high-privacy immersion is impossible.

3.3 Ambient Connection vs. Environmental Conflict

While the active app usage caused friction, the passive ambient features (light and music) bridged the gap between contexts. P4 and P7 noted that the mood history in the App and light colors shown on the coaster served as effective *"topic-openers"* for starting conversations with others, especially with strangers or guests. P2 and P3 found that the colors of lights helped them ask friends about their feelings. P5 envisioned the coaster as a data-driven "*ice-breaking intermediary"* for bartenders to customize service.

Again, professionals foresaw challenges in future deployments. P1 (Bartender) and P5 warned of "sensory interference," noting that the coaster's light and music could conflict with the bar's existing atmosphere (background music and lighting). This indicates that slow technology cannot simply overlay an existing environment but must respect its sensory hierarchy.

4 DISCUSSION

Current research on slow technology and design often frames deceleration and friction as universally positive attributes that foster mindfulness and meaning-making (Cox et al., 2016; Grosse-Hering et al., 2013; Marsh, 2016). However, our situated investigation reveals a critical limitation in this view: the same "slowness" that facilitates healing and reflection in solitude is often perceived as a social "interruption" in group settings. This suggests that the value of slow interaction is not inherent to the device but is contingent upon the social context. We propose that successful slow design must move beyond static friction toward dynamic calibration, guided by three design insights. We also identified one research gap in digital humor design.

4.1 Gentle Revelation: Prioritizing "Calm" over "Immersion "

Traditional interactive systems often compete for user attention through direct notifications (vibrations, sounds). While slow design advocates for "rich" experiences (Strauss & Fuad-Luke, 2008), our findings in the bar context highlight a risk of sensory overload. Participants rejected high-fidelity feedback (music and screen interactions) because it conflicted with the bar's existing auditory landscape.

Instead, we propose the strategy of Gentle Revelation. Drawing on Weiser's principles of calm technology (Weiser & Brown, 1996, 1997), slow interfaces in social spaces should not demand cognitive processing but rather invite it through "peripheral cues." Our participants' preference for the ambient light—which could be ignored or attended to at will—demonstrates this principle. Unlike "immersive" systems that sequester the user, a "gentle" system unfolds information gradually (Weiser & Brown, 1996), ensuring the artifact supports the social atmosphere rather than hijacking the conversation.

4.2 Reflective Pacing: From "Explicit" to "Opt-in" Slowness

How should friction be implemented? Odom (2024) distinguishes between Explicit Slowness (where the system restricts control to force a pause, e.g., the Photobox (Odom et al., 2010)) and Implicit Slowness (where the system invites reflection but remains open). Our study provides empirical evidence that Explicit Slowness is fundamentally incompatible with social flow. When our system enforced a "loading delay" to induce mindfulness, users in the bar felt it was a "bug" or an awkward pause that severed eye contact with friends.

Therefore, we argue that socially situated devices must utilize Opt-in Friction. The artifact should default to a "Low-Friction Ambient" state (e.g., subtle lighting) to preserve the social rhythm, transforming into a "High-Friction Ritual" state only when the user creates a "micro-boundary" by actively engaging (Cox et al., 2016)—such as physically scanning the coaster. This resolves the tension identified in our findings: users desire the option for deep engagement (Ritual) without the intrusion of enforced delays (Interruption) (Sengers et al., 2005).

4.3 Contextual Interaction Friction: Leveraging Social Awareness

Finally, we challenge the binary view of "fast" vs. "slow" design. Most context-aware computing utilizes sensor data to eliminate friction and predict user needs (Dey, 2001). We argue for an inverse approach: leveraging context to strategically insert friction. In solitary contexts, the system should increase interaction friction (e.g., by requiring tangible inputs or mood logging). Here, the delay is not inefficiency but the "value proposition" itself, creating headspace for self-dialogue (Sengers et al., 2005). In group settings, the system must shift to seamless operation; features such as "social history" should serve only as passive icebreakers (Weiser & Brown, 1997) rather than as active tasks.

This implies that "slowness" must be socially context-aware (Hallnäs & Redström, 2001). Future slow technology should not only sense where the user is but whom they are with, dynamically matching the effort it demands to the social etiquette of the moment.

4.4 Physical Symbolism and the Limits of Digital Humor

The original Kansei Engineering analysis identified Humor as a key emotional attribute in the design space, informed by benchmark products—such as zoomorphic coasters—whose symbolic physical forms directly trigger meaning at the "Experience of Meaning" level (Karana, Barati, et al., 2015). By contrast, the handcrafted glass prototype operates primarily at the visceral level, offloading the "unexpected" dimension onto digital feedback: AI-curated music recommendations and ambient light responses.

Within a single 20-minute session, these digital surprises proved insufficient to establish the depth of emotional connection or genuine humor response that strong physical metaphor can achieve more immediately. This exposes a cross-medium design gap: static physical symbolism encodes humor durably through material form, whereas dynamic digital feedback requires repeated exposure to accumulate equivalent meaning. A productive direction for future Kansei tangible computing research lies in how to more deeply integrate the dynamic responsiveness of AI software with the durable symbolic resonance of physical hardware, achieving coherent humor across both levels of experience.

5 CONCLUSIONS AND LIMITATIONS

This study employed a three-phase Research through Design methodology to investigate how slow design principles can be embedded in a familiar daily object. In Phase I, a Kansei Engineering Type I approach (a Semantic Differential survey; 15 samples, N=32) yielded three factors via Exploratory Factor Analysis (explaining 88% of the variance) and revealed an unmet design opportunity: artifacts that combine emotional richness, high interactivity, and humor remain rare among existing coasters. Guided by this gap, Phase II produced the *Slow Coaster*: an NFC-enabled, handcrafted glass coaster with embedded ambient LED feedback and a web application that supports mood logging, AI-curated music, and a shared social history.

Phase III evaluated the prototype in a commercial bar (N=5) and a domestic setting (N=3) using a Wizard-of-Oz protocol. Analysis revealed three themes: *Solitary Resonance* (high-friction features fostered mindful "healing" rituals in quiet contexts); *Social Dissonance* (the same features disrupted the flow of group conversation); and *Ambient Connection* (passive ambient cues served as low-friction conversation starters across both contexts). From these findings, we propose three design provocations: *Gentle Revelation* (using peripheral ambient cues rather than attention-demanding notifications), *Reflective Pacing* (defaulting to low-friction ambient states with opt-in ritual engagement), and *Contextual Interaction Friction* (dynamically calibrating interaction effort to social etiquette). Together, they argue that slowness must be socially calibrated rather than universally prescribed.

Several limitations bound these contributions. The AI music recommendation and emotional mirroring light were simulated using predefined playlists and a Wizard-of-Oz protocol, which constrained claims about long-term responsiveness and personalization; future work should deploy a fully functional AI backend to evaluate whether genuine dynamic personalization deepens reflective and humorous qualities over time. The social history feature relied entirely on pre-built data so that a live deployment will require explicit consent mechanisms and careful privacy design around data ownership and opt-in transparency. Finally, the small sample size (eight participants across two contexts) limits the generalizability of these findings and captures *ritual initiation* rather than ritual formation; longitudinal deployment with a larger and more diverse sample is necessary to examine whether initially disruptive friction is reframed as ritualistic through habituation, and whether the proposed provocations generalize across broader social configurations.

ACKNOWLEDGMENTS

This research was supported by the National Science and Technology Council of Taiwan (NSTC114-2410-H-007-081).

REFERENCES

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291. https://doi.org/10.1016/j.ijhcs.2006.11.016

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chapman, J. (2012). *Emotionally durable design: Objects, experiences and empathy*. Routledge.

Chuang, Y. (2020). Designing the expressivity of multiple smart things for intuitive and unobtrusive interactions. In *Proceedings of the conference on designing interactive systems* (pp. 2007-2019). ACM. https://doi.org/10.1145/3357236.3395450

Cox, A. L., Gould, S. J., Cecchinato, M. E., Iacovides, I., & Renfree, I. (2016). Design frictions for mindful interactions: The case for microboundaries. In *Proceedings of the SIGCHI conference on human factors in computing systems* (Extended abstracts, pp. 1389-1397). ACM. https://doi.org/10.1145/2851581.2892410

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, *5*(1), 4–7. https://doi.org/10.1007/s007790170019

Fuad-Luke, A. (2002, December 2). *Slow design: A paradigm shift in design philosophy*. Paper presented at the 2nd international conference on open collaborative design for sustainable, Bangalore, India.

Grosse-Hering, B., Mason, J., Aliakseyeu, D., Bakker, C., & Desmet, P. (2013). Slow design for meaningful interactions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3431-3440). ACM. https://doi.org/10.1145/2470654.2466472

Hallnäs, L., & Redström, J. (2001). Slow technology–designing for reflection. *Personal and Ubiquitous Computing*, *5*(3), 201–212. https://doi.org/10.1007/PL00000019

Hassenzahl, M. (2010). *Experience design: Technology for all the right reasons*. Springer. https://doi.org/10.1007/978-3-031-02191-6

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products–Facets of user experience. *Interacting with Computers*, *22*(5), 353-362. https://doi.org/10.1016/j.intcom.2010.04.002

Johnson, K. W., Lenau, T., & Ashby, M. F. (2003). The aesthetic and perceived attributes of products. In *Proceedings of the 14th international conference on engineering design* (pp. 173-174). Design Society.

Karana, E., Barati, B., Rognoli, V., & Zeeuw Van Der Laan, A. (2015). Material driven design (MDD): A method to design for material experiences. *International Journal of Design*, *9*(2), 35–54.

Karana, E., Pedgley, O., & Rognoli, V. (2015). On materials experience. *Design Issues*, *31*(3), 16–27. https://www.jstor.org/stable/43829331

Marsh, T. (2016). Slow serious games, interactions and play: Designing for positive and serious experience and reflection. *Entertainment Computing*, *14*, 45–53.

Norman, D. A. (2004). Introduction to this special section on beauty, goodness, and usability. *Human–Computer Interaction*, *19*(4), 311-318. https://doi.org/10.1207/s15327051hci1904_1

Norman, D. A. (2007). *Emotional design: Why we love (or hate) everyday things*. Basic Books.

Odom, W. (2024). Illustrating, annotating & extending design qualities of slow technology. In *Proceedings of the NordicCHI conference on human-computer interaction* (Adjunct, Article No.74). ACM. https://doi.org/10.1145/3677045.3685499

Odom, W., Banks, R., & Durrant, A. (2010). Designing for slowness, anticipation and re-visitation: A long term field study of the photobox. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1961-1970). ACM. https://doi.org/10.1145/2556288.255717

Odom, W., Stolterman, E., & Chen, A. Y. S. (2022). Extending a theory of slow technology for design through artifact analysis. *Human–Computer Interaction*, *37*(2), 150-179. https://doi.org/10.1080/07370024.2021.1913416

Osgood, C. E. (1957). *The measurement of meaning*. University of Illinois.

Pallasmaa, J. (2024). *The eyes of the skin: Architecture and the senses*. John Wiley & Sons.

Schütte, S. (2005). *Engineering emotional values in product design—Kansei engineering in development* [Unpublished doctoral dissertation]. Linköping University, Sweden.

Sengers, P., Boehner, K., David, S., & Kaye, J. J. (2005). Reflective design. In *Proceedings of the 4th decennial conference on critical computing* (pp. 49-58). ACM. https://doi.org/10.1145/1094562.1094569

Strauss, C., & Fuad-Luke, A. (2008, July 12). *The slow design principles*. Paper presented at the conference on changing the change, Politecnico di Milano and Politecnico di Torino, Turin, Italy.

Weiser, M., & Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal*, *1*(1), 75-85.

Weiser, M., & Brown, J. S. (1997). The coming age of calm technology. In *Beyond calculation: The next fifty years of computing* (pp. 75-85). Springer.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach (pp. 310-319). In *Proceedings of the 8th conference on designing interactive systems*. ACM. https://doi.org/10.1145/1858171.1858228

APPENDIX

Table A1 presents the final 10 bipolar adjective pairs selected in this study, which were operationalized across five of the six slow design principles.

Bipolar Adjective Pairs	Mapped Slow Design Principle	Rationale for Mapping
Low-tech vs. High-tech	Reveal	Low-tech mechanisms transparently uncover everyday processes that are typically obscured by complex digital integration.
Handmade vs. Manufactured	Reveal	Traces of manual crafting explicitly reveal the resources, time, and human labor invested in the creation process.
Cheap vs. Expensive	Reveal	The perceived cost or exclusivity directly reveals underlying socioeconomic implications and material value.
Emotional vs. Functional	Reflect	Emotional designs foster deep affective connections, prompting users to introspect on their personal values and memories.
Novel vs. Ordinary	Reflect	Unconventional forms disrupt automatic usage patterns, forcing users to pause and reflect on their behaviors.
Low Interactive vs. High Interactive	Engage	High interactivity (e.g., intentional friction) demands cognitive and physical effort, encouraging active engagement over passive consumption.
Monotonous vs. Rich	Engage	A high density of visual or sensory details requires active exploration, preventing the user from processing the object passively.
Humorous vs. Humorless	Expand	Playful and humorous elements subvert expectations, extending the design's implications into broader cultural and social contexts.
Classic vs. Trendy	Evolve	Classic designs embody historical permanence, allowing the object's meaning to mature and develop over time.
Durable vs. Disposable	Evolve	Physical longevity provides the necessary material foundation for an object to age sustainably and evolve alongside the user.

DESIGNING WHERE IT FEELS DEEPEST

MAPPING EMOTIONAL VALENCE TO IMPROVE HUMAN PERFORMANCE

Rachael VOLKER[1a]

[a] *Liverpool School of Art and Design, Liverpool John Moores University, Liverpool, United Kingdom, rachaelvolker@proton.me*

ABSTRACT

Design has always shaped how people feel. Yet in modern product creation, the emotional consequences of design decisions are rarely measured. They are inferred, assumed, or set aside. This paper argues that overlooking emotion is not only a design concern but a performance problem, and that the cost of getting it wrong is often invisible precisely when it matters most.

When a product creates friction, whether a seam that irritates, a fabric that traps heat, or a garment that fails in public, the cost is rarely just physical discomfort. It fractures attention, recruits effort the body would rather spend elsewhere, and quietly erodes the confidence that performance depends on. Conversely, when design gets it right, the product disappears and the user gets their focus back. That is not a soft benefit, it is a measurable shift in human capacity.

We frame these effects through the lens of compensatory control (Hockey, 1997). Under stress or load, people protect their primary goal by recruiting additional effort, so performance can look stable on the surface while the system pays a hidden price in physiological strain, narrowed attention, and downstream fatigue. Hockey calls this a latent decrement. A product that forces this trade-off does not simply annoy the athlete; it taxes the very system the product is meant to support.

This paper introduces the Convergent Affective Design Evaluation (CADE) framework, a multimodal methodology that locates these moments by triangulating across physiological, vocal, and behavioral indicators of emotion. Rather than relying on any single signal, CADE aligns multiple channels of the body's response, including heart rate variability, vocal prosody, and other measurable physiological inputs, into a single time-aligned composite of emotional experience, identifying not just whether a user liked something, but when a design element began to make or break their capacity to perform. Initial athlete testing shows that seemingly minor features, such as ventilation placement or seam construction, can produce rapid, measurable shifts in emotional and physiological state.

[1] Corresponding author.

The contribution is both methodological and philosophical: designing for emotional impact is a core performance intervention. This paper offers a practical framework that translates emotion into analyzable evidence, helping teams find the design opportunities that reduce load, protect confidence, and sustain endurance. Ultimately, the argument is simple. Designing for performance requires designing for feeling, and feeling can be mapped, studied, and used.

Keywords: emotional design; emotional valence; compensatory control; product design; affective neuroscience

1 INTRODUCTION

This research explores how the overlap of emotional response and physiological data can inform the way we design products, especially products worn on the body. Prior work in product-experience research shows that affect is not peripheral; it is tightly linked to how people evaluate products, form attachment, and continue or discontinue using them over time (Schifferstein & Zwartkruis-Pelgrim, 2008). This paper introduces a framework for identifying where design feels deepest by isolating the features that elicit the strongest emotional and physiological responses. When those responses are supportive, people move, focus, and endure. When they are not, performance quietly collapses, often without leaving an obvious trace.

The central move is to treat that quiet collapse as something we can detect. Psychophysiological indicators such as cardiac activity have been shown to meaningfully model user-experience states, offering a complementary layer to self-report when assessing stress, engagement, and interaction quality (Čertický et al., 2019). By capturing both what users feel and how their bodies respond to products, we can locate opportunities for innovation that genuinely tend to human needs. This methodology turns how it feels into usable evidence, allowing teams to pinpoint the exact moments a product supports performance or undermines it. When those moments are mapped to specific features, designers can make targeted changes that reduce cognitive load, protect confidence, and sustain physical endurance.

We invite the reader to hold a larger possibility in mind throughout. If we can reliably see where a product helps a body settle and where it forces a body to fight, we are no longer only refining objects. We are designing for the conditions under which people can be fully present in their own effort, and that is a horizon worth reaching for.

2 CONTEXT AND BACKGROUND

Design has always been intertwined with emotion. People develop deep attachments to certain products, particularly those they keep in extended contact with, such as apparel, footwear, and furniture. These products can spark joy and security or trigger frustration, sadness, or fear (Schifferstein & Spence, 2008). Emotional design theory holds that products engage us at visceral, behavioral, and reflective levels, and that fulfilling experiences across all three foster positive emotion (Norman, 2004). Movements such as Positive Design extend this further, arguing that design can be a deliberate instrument for human flourishing and well-being (Desmet & Pohlmeyer, 2013).

Physical performance depends on a body that can breathe and move without constant negotiation. Mental performance depends on attention that is not fractured by irritation, self-consciousness, or threat. When a product repeatedly creates moments of stress, it does not merely irritate the user. It taxes the nervous system, disrupts focus, and can shape whether a person returns to the activity at all. Prior research in design psychology shows that strong product attachments are dominated by positive emotion (Schifferstein, 2008), while negative experiences, especially discomfort, product failure, and public embarrassment, can leave a lasting emotional residue that undermines trust and confidence.

2.1 A mechanism for the hidden cost: compensatory control

To explain why a small design flaw can carry an outsized cost, we draw on Hockey's cognitive-energetical framework of compensatory control (Hockey, 1997). Developed to account for human performance under stress and high workload, the model makes a deceptively simple observation: primary task performance is remarkably stable under stress, and that stability comes at a cost, elevated physiological activation and subjective strain. People protect what matters most by mobilizing effort, and that effort has a price.

Crucially, the price is often hidden. Hockey argues that even when overt performance holds, the strain of protecting it surfaces elsewhere, as the neglect of secondary tasks, a shift to simpler and less flexible strategies, increased sympathetic activation, and fatigue after the fact. He names this a latent decrement: a compromised internal state that stays invisible under normal conditions but constrains the system's capacity to absorb anything further. The model also distinguishes the modes by which people cope. In active coping, demands are met with effort but little distress. In a strain mode, performance is maintained, but at a steep and sustained physiological cost. In passive coping, the goal itself is downgraded or abandoned. The shift that is concerning is the slide from strain into disengagement, where the athlete, worn down by accumulated cost, simply stops showing up.

Reframed this way, a poorly placed seam or a constricting, heat-trapping garment is exactly the kind of disturbance Hockey describes: an external perturbation that pulls resources away from the task. The athlete keeps running, so her split time may look untouched. But she is now paying a compensatory cost, and that cost is the very thing conventional product feedback fails to capture. Our method is, in effect, an instrument for making latent decrements visible in the lived experience of a product.

2.2 Reading the body: heart rate variability as a window on regulation

If compensatory effort leaves a physiological signature, heart rate variability (HRV) is one of the most informative places to look for it. The neurovisceral integration model (Thayer et al., 2012) describes a flexible cortical-subcortical network, centered on the ventromedial prefrontal cortex and amygdala, that exerts top-down inhibitory control over the heart. When that control is intact, the heart shows rich beat-to-beat variability, the marker of a system that is regulated, adaptable, and at ease. When perceived threat draws prefrontal regions offline, this inhibition is released, sympathoexcitatory circuits (the stress pathways that mobilize energy for action) are disinhibited, and HRV falls. This gives our physiological measure a clear interpretive frame. A drop in HRV during a product experience is not noise; it is a reading on how hard the regulatory system is working, and a sign that the body has shifted toward a more stress-loaded, less flexible state.

Thayer's account dovetails with Hockey's: both describe a system that protects its goals through top-down control, and both warn that when that protection is sustained, the unrelieved load accumulates, an idea Thayer connects to allostatic load and long-term wear on the body.

3 MATERIALS AND METHODS

Eight women took part in the study. For this paper, the focus was on session in which participants recounted clearly differentiated product experiences, distinct moments of both distress and relief, since these contrasts are precisely where a convergence-based method can be most clearly demonstrated. The interview presented below is the single clearest case of this pattern offered as an illustrative example rather than a representative average.

This research uses a multimodal approach to capture users' emotional responses to products in a controlled but familiar environment. Participants interacted with, or recalled experiences of, products while their narrative, vocal expression, and cardiac activity were recorded together. The design deliberately combines a subjective channel (what the participant says and how they say it) with an objective physiological channel (how the body responds), so that each can be read against the other.

Cardiac activity was recorded with a Polar H10 ECG chest strap, from which heart rate variability was derived as an index of autonomic regulation (Thayer et al., 2012). Emotion is reliably accompanied by changes in cardiac and vocal activity (Čertický et al., 2019; Kreibig, 2010), and the voice in particular carries a rich, involuntary trace of affective state. Vocal expression was analyzed acoustically in Praat (Boersma & Weenink, 2024), extracting prosodic features such as fundamental frequency and its variability, intensity, and measures of voice quality (for example jitter and shimmer) that shift with arousal and emotional strain. Interview narrative was transcribed and coded so that specific product moments could be located on a common timeline.

The core methodological shift is simple but consequential: build a single, time-aligned dataset that overlays emotional indicators with physiological response. The result is not a list of reported feedback but a map of emotional magnitude across the product experience. This builds on established practice in human-computer interaction, where psychophysiological signals are combined with subjective report to model user-experience states (Čertický et al., 2019). Applied to design research, the framework supports a more objective synthesis of emotional response by linking what people say to measurable, time-stamped changes in the

body (Figure 1). Importantly, no single metric or method is relied upon for interpretation alone.

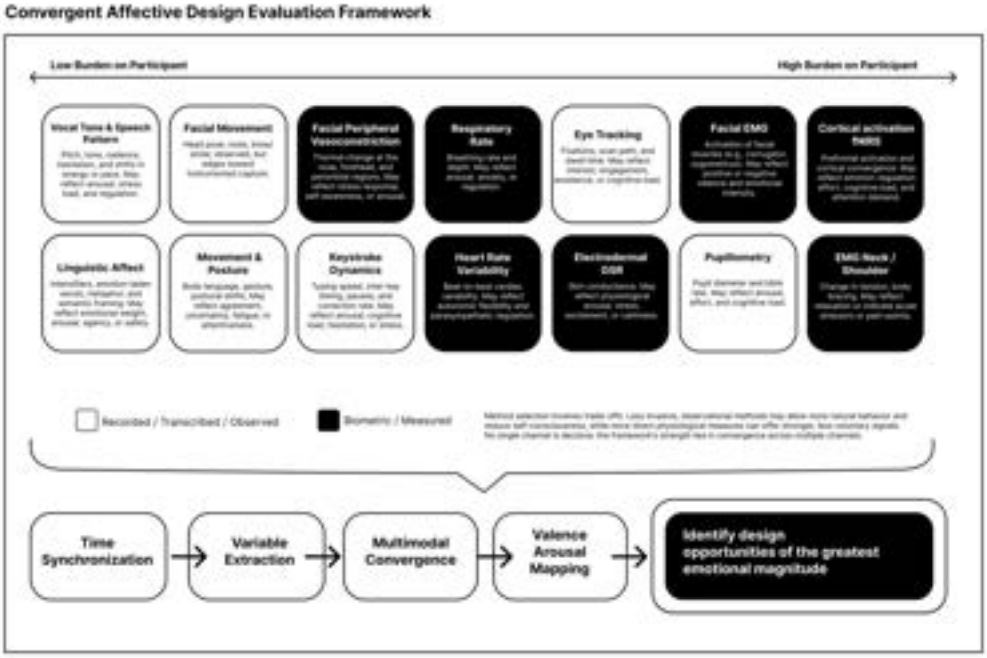


Figure 1. ***The Convergent Affective Design Evaluation (CADE) framework.*** *CADE locates the moments in an experience where a design carries the greatest emotional and performance consequence. It draws on a broad, substitutable menu of affective signals, arranged (upper panel) from unobtrusive, observed signals to more obtrusive, directly measured physiological ones; the signals shown are illustrative, not prescriptive. No single signal is treated as decisive. Instead, convergence is the core mechanism: when independent signals agree at the same moment, that agreement gives the moment its weight. The lower panel shows the pipeline, time synchronization, feature extraction, multimodal convergence, and valence arousal mapping, culminating in the design opportunities of greatest emotional magnitude.*

4 KEY FINDINGS AND INSIGHTS

Initial findings show clear patterns in how design elements trigger emotional and physiological response. Even when participants were recalling a single product experience rather than living it in the moment, they exhibited a wide range of affect. When a product caused discomfort or failed, and particularly when it failed in public, participants' accounts carried the markers of fear and anger, and these moments coincided with pronounced shifts in cardiac activity.

One insight stands out: seemingly minor design issues can provoke outsized reactions. A poorly placed seam that irritates does not only create physical discomfort. It creates a distraction that fractures focus and temporarily narrows cognitive capacity, the textbook signature of a latent decrement. From a performance standpoint that is far from trivial. It is the difference between being fully in it and being repeatedly pulled out.

4.1 Multimodal emotional response in an athlete interview

In one interview, a participant described a range of pain points across sports bras, from irritation to fit to heat management. Read as a transcript alone, many of these concerns could have been weighted as roughly equal. The multimodal data told a different story. When she described feeling trapped in a sweaty sports bra she hated (Bra A), her language was direct, but

the physiological and vocal signals were disproportionate to the rest of the conversation. As she spoke about the constriction she felt and the sweat-logged fabric, her vocal-acoustic profile shifted in ways consistent with elevated arousal and strain, while her heart rate variability dropped to 24ms.

Read through the neurovisceral lens, that HRV reading is a window onto autonomic regulation. A drop of this kind reflects reduced parasympathetic flexibility and a shift toward a more stress-loaded state, especially when it occurs acutely during the recall of a specific event (Thayer et al., 2012). It suggests this was not a passing annoyance but a moment that meaningfully engaged threat and a sense of lost control, an interpretation reinforced rather than asserted, because the vocal signal and the narrative pointed the same way at the same instant (Table 1).

Table 1. Vocal-acoustic and cardiac comparison, Bra A versus Bra B

Measure	Baseline 00;00;00;00 - 00;10;04	Bra A (distress) 00;00;05;00 - 00;05;35;08	Bra B (relief) 00:00:53:22 - 00:01:27:22
F0 variability (SD, Hz)	38.2	48.48	29.86
Jitter, local (%)	1.81	2.07	1.62
Shimmer, local (%)	13.1	15.38	11.26
Harmonics-to-noise ratio (dB)	11.8	7.93	16.75
Heart rate variability (RMSSD, ms)	33	24	41

Note: Voice-quality and pitch measures were extracted in Praat (Boersma & Weenink, 2024) using autocorrelation pitch tracking with a 75–500 Hz analysis range; jitter, shimmer, and harmonics-to-noise ratio are from the Praat Voice Report. Intensity is not reported because the recording was not calibrated for absolute sound pressure level. The fraction of unvoiced frames was comparable across segments (A = 46.4%, B = 47.7%), indicating that differences in voice quality are not attributable to one segment carrying a noisier signal. Heart rate variability (RMSSD) was derived from Polar H10 ECG over the corresponding segment.

Each session was video-recorded and reviewed to confirm that changes in prosody and heart rate variability were not attributable to shifts in posture, body position, or other movement-related factors. Participants remained in a consistent seated position throughout, supporting an affective rather than postural interpretation of the observed differences. To protect participant privacy, this footage was used solely for internal verification and is not reproduced here.

A baseline was also captured for each participant at the start of the session, while they read an emotionally neutral, deliberately bland control script. This baseline provides a within-participant reference point against which the distress and relief segments can be compared, helping to distinguish emotionally driven change from each individual's ordinary speaking and cardiac signature.

Later in the same interview, she spoke about a sports bra that finally relieved the pain point she had described (Bra B). As she described generous front perforations that let air reach the sternum and chest while running, her vocal cadence softened and her account turned warm and unhurried, the signature of a body that had downshifted out of threat. The implication is not that one bra is universally good and another bad, but that specific design features can rapidly change emotional and physiological state in ways that are measurable and actionable.

This is where the methodology becomes performance-relevant. Had the interview been evaluated on transcript alone, the participant's list of pain points might well have been treated as equal in weight. Without clear prioritization, a design and development team would likely have spread incremental improvements thinly across many features, or addressed only those the budget and timeline allowed. By layering cardiac response and vocal-acoustic analysis onto the narrative, we can identify the single most important opportunity with far greater confidence. Here the data surfaced the priority unambiguously: the sweat-trapped, constricting experience was the dominant driver of distress, and therefore the most impactful opportunity to pursue. Across the session, the most intense moments were consistently the ones that stood out across both physiological and self-report channels, helping translate lived experience into clear design direction. The same convergence pattern, physiological and vocal signals intensifying together at moments of narrated distress, recurred across the differentiated sessions, indicating that this interview illustrates a repeated dynamic rather than an isolated instance.

4.2 Implications

These findings show that when traditional design research is paired with a dimensional view of emotion and grounded in physiological signal, designers gain a far sharper view of what truly matters in lived experience. Self-report remains essential, but on its own it can flatten nuance, making many pain points sound equally urgent while masking which moments are actually driving distress, distraction, or withdrawal. Combining interviews with physiological indicators and brief valence-focused tools (for example the Self-Assessment Manikin; Bradley & Lang, 1994) lets us identify the highest-leverage opportunities for innovation, the places where a single feature can meaningfully restore comfort, confidence, and control.

In performance contexts, that clarity is a practical advantage. It helps teams prioritize the changes that protect focus, reduce strain, and sustain endurance, especially in high-exertion situations where body and mind are already operating near their limits. In Hockey's terms, it tells us where a product is forcing an athlete into a costly strain mode, and where a better design lets her return to low-cost, active engagement, the state in which she can give the most while losing the least.

5 CONCLUSION

Design decisions leave traces in the body. In this work, those traces become legible. By synchronizing interview narrative with cardiac and vocal-acoustic indicators, we move from generalized preference statements toward a precise understanding of when a product supports human capacity and when it quietly drains it. The central finding is not simply that products evoke emotion, but that the most consequential design problems are often the ones a transcript

alone would never elevate. When emotional valence and physiological strain converge on a single feature, the design priority becomes unmistakable. This gives designers a practical, repeatable way to protect what matters most in performance: attention, confidence, and the will to persist.

The implications reach well beyond sport. In a broader landscape where stress, anxiety, and cognitive fatigue are increasingly common, the everyday products carried through daily life have real power to either amplify strain or restore regulation. A drop in HRV during a product experience marks a moment of diminished autonomic flexibility, a body shifting into a more stress-loaded state; left unaddressed, such micro-stressors may accumulate into meaningful physiological burden over time (Thayer et al., 2012). Conversely, when design reduces irritation, improves comfort, or removes the threat of failure, the nervous system can settle and the user regains bandwidth for the task at hand. Affective experience, as a growing body of work suggests, is not a cosmetic layer of interaction but a pathway through which the things around us shape well-being (Kreibig, 2010; Shu et al., 2018).

So it is worth dreaming a little about where this leads. Imagine product testing that listens to the body as carefully as it listens to opinion. Imagine design teams who can see, in advance, which detail or material or feature will let an athlete stay present in her own effort, and which will quietly pull her out of it. Imagine performance products evaluated not only by what they let people do, but by how steady they let people feel while doing it. This is the larger invitation of the work: to treat the body's quiet signals as design intelligence, and to build things that give people their attention back.

Ultimately, this paper argues for a redefinition of performance design. Performance is not only strength, speed, or endurance. It is also the ability to stay steady inside the experience, to hold attention, to feel safe enough in one's own body to keep going. The methodology presented here offers a way to design toward those outcomes with greater clarity, accountability, and compassion. It is a call to treat how it feels not as a subjective footnote, but as evidence, and as a primary lever for improving human performance in the moments that matter most.

6 ACKNOWLEDGMENTS

This research emerges directly from the author's doctoral work in the Liverpool School of Art and Design at Liverpool John Moores University (LJMU), where her PhD examines neuroaesthetics and the emotional and physiological dimensions of design experience. The theoretical foundations, the CADE framework, and the methodological approach presented in this paper developed through that broader program of work, which sits at the intersection of product design, design research, and the psychophysiology of human experience investigating how designed objects shape emotion, physiology, and performance. This study reflects that ongoing research and was carried out through the author's own practice. The author also

serves as Adjunct Faculty in the College of Design at the University of Oregon, teaching design research, sports product design, and innovation.

The author thanks the participants who generously shared their time and experiences. All participants took part voluntarily and gave informed consent for the recording of vocal and physiological (heart rate variability) data, which were handled confidentially and used solely for the purposes of this research.

7 REFERENCES

Boersma, P., & Weenink, D. (2024). Praat: Doing phonetics by computer [Computer program]. https://www.praat.org

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Čertický, M., Čertický, M., Sinčák, P., Magyar, G., Vaščák, J., & Cavallo, F. (2019). Psychophysiological indicators for modeling user experience in interactive digital entertainment. Sensors, 19(5), 989. https://doi.org/10.3390/s19050989

Desmet, P. M. A., & Pohlmeyer, A. E. (2013). Positive design: An introduction to design for subjective well-being. International Journal of Design, 7(3), 5–19. https://www.ijdesign.org/index.php/IJDesign/article/view/1666

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. Biological Psychology, 45(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books.

Schifferstein, H. N. J., & Spence, C. (2008). Multisensory product experience. In H. N. J. Schifferstein & P. Hekkert (Eds.), Product experience (pp. 133–161). Elsevier. https://doi.org/10.1016/B978-0-08-045089-6.50008-3

Schifferstein, H. N. J., & Zwartkruis-Pelgrim, E. P. H. (2008). Consumer-product attachment: Measurement and design implications. International Journal of Design, 2(3), 1–14. https://www.ijdesign.org/index.php/IJDesign/article/view/325

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A review of emotion recognition using physiological signals. Sensors, 18(7), 2074. https://doi.org/10.3390/s18072074

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. Neuroscience & Biobehavioral Reviews, 36(2), 747–756. https://doi.org/10.1016/j.neubiorev.2011.11.009

DEVELOPMENT OF A MIXED REALITY–BASED ORGAN REGISTRATION ASSISTANCE SYSTEM FOR CT-GUIDED PUNCTURE

Kotone.T [a],Motoki.N [b],Ikuto.T [b],Izumi.H [a],Joue.C [a]

[a] *Kogakuin University, Japan, j222242@g.kogakuin.jp*
[b] *Tokyo Medical College, Japan*

ABSTRACT

Extended Reality (XR) technologies have been increasingly applied in the medical field, particularly for supporting image-guided interventions. In CT-guided biopsy, accurate puncture requires repeated CT imaging to confirm the insertion path, which results in increased radiation exposure for both patients and operators. Mixed Reality (MR) has the potential to reduce this exposure by overlaying virtual anatomical information onto physical space; however, conventional MR-based alignment methods suffer from positional errors and limited operability when registering virtual objects to real organs.

In this study, we propose an MR-based alignment method that reduces registration error by introducing two-dimensional positional markers overlaid on physical magnetic markers. The proposed approach aims to improve both alignment accuracy and usability when superimposing virtual objects onto real-world organ positions. An MR application was developed using Meta Quest 3 and Unity, enabling marker acquisition and virtual object placement through a combination of positional marker detection and controller-based interaction.

To evaluate the effectiveness of the proposed method, we conducted a user study with 22 participants and compared two approaches: (1) visual manual placement of virtual markers and (2) automatic placement assisted by positional markers. Alignment accuracy, operation time, and subjective usability were assessed. Results showed that the proposed positional marker–assisted method significantly reduced positional error in the X and Y directions and shortened operation time compared to the visual placement method ($p < 0.05$). Subjective evaluations also indicated improved ease of operation and visual naturalness.

These findings demonstrate that the proposed method effectively improves registration accuracy and usability in MR-based alignment for CT-guided biopsy support, suggesting its potential to contribute to safer and more efficient image-guided interventions.

Keywords: *MR, Biopsy, Registration*

1 INTRODUCTION

In recent years, cross-reality technologies that integrate the real world and virtual environments—collectively referred to as Extended Reality (XR)—have attracted increasing attention across a wide range of fields. The characteristics of XR technology as a medium include three-dimensional information presentation and the extension of time and space, thereby enhancing the user experience (UX). These characteristics support more intuitive information understanding and spatial perception. In the medical field, enabling clinicians to intuitively understand spatial relationships and procedural information can reduce psychological burden and anxiety, while improving their sense of confidence and overall satisfaction with medical practice. In particular, in the medical domain, both research and practical applications utilizing XR technologies, such as surgical simulation and rehabilitation support, have been actively developed [1][2][3][4].

In the diagnosis of tumor lesions in clinical practice, when definitive diagnosis cannot be achieved through imaging modalities alone, biopsy plays a crucial role in determining treatment strategies. In CT-guided biopsy procedures, a safe puncture path is planned based on CT images, and the needle is inserted accordingly. However, securing an appropriate insertion route often requires multiple CT scans, leading to increased radiation exposure for both patients and operators. Therefore, the development of techniques to reduce radiation dose has become an important challenge.

Previous studies have demonstrated the potential of using Mixed Reality (MR) to support CT-guided puncture by acquiring and visualizing needle position and insertion trajectory information in real time. Nevertheless, issues related to operability and insufficient confirmation of the actual organ position have been reported as limitations [2]. To address these challenges, our preliminary study proposed a method using multiple magnetic markers that can be attached to a three-dimensional human body surface and detected by CT scanning. By calculating the centroid of the marker coordinates, a virtual object could be overlaid onto the physical object, and the method was reported to be both effective and easy to operate [3]. However, in this approach, the physical markers must first be detected by the application to generate corresponding coordinates in virtual space before alignment, and errors are introduced during this process.

In this study, we aim to reduce the positional discrepancy between “markers generated in virtual space based on detected physical markers” and the “actual physical markers” when overlaying virtual objects onto real organ positions using MR. Instead of relying on visually guided marker placement, we introduce additional two-dimensional positional markers to accurately acquire the positions of physical magnetic markers. The goal of this research is to develop an MR-based content that enables accurate marker acquisition while maintaining ease of operation.

2 APPLICATION DEVELOPMENT ENVIRONMENT AND OVERVIEW

2.1 Development Environment

In this study, the head-mounted display (HMD)-type MR device Meta Quest 3 was used for application implementation. Unity was adopted as the integrated development platform, and Visual Studio was used as the development environment.

2.2 Application Overview

The overlay function of the developed application and the marker acquisition process are described below.

Step 1. Four magnetic markers (reference points) with a diameter of 2.2 cm (magnet diameter: 0.8 cm) and a height of 0.25 cm, which are detectable by CT scanning (Fig. 2), are placed on a physical object (Fig. 1: an abdominal biopsy phantom consisting of a membrane simulating skin and fat layers, containing a gel-encapsulated tumor-mimicking target; hereafter referred to as the phantom). A virtual object is generated based on the data obtained from CT scans.

Step 2. Using the four reference points of the virtual object, the centroid is calculated, and this point is defined as the origin of the virtual object's coordinate system.

Step 3. The four magnetic markers (reference points) on the physical object are acquired, and the physical coordinates are generated in the virtual space within the application. At this stage, the center of each reference point is treated as the center of a sphere, and spherical virtual markers are placed accordingly.

Step 4. By overlaying the coordinate system of the virtual object onto the generated physical coordinates, the virtual object is positioned as shown in Fig. 3.

Method ①: Visual Marker Placement

In Step 3, a controller is used to visually align and place the virtual markers so that they overlap with the physical markers, as illustrated in Fig. 4.

Method ②: Positional Marker–Assisted Placement

To reduce the error associated with visual placement in Method ①, this study introduces a square positional marker with side length 2.2 cm, which is attached on top of each magnetic marker (diameter: 2.2 cm), forming a dual-marker configuration (Fig. 5). Using QuestCameraTools-Unity, a Unity-based alignment library available for Meta Quest 3, the application was implemented to display virtual markers directly on the positional markers.

In Step 3, identification numbers were assigned to the markers as shown in Fig. 6. A laser beam was generated from the controller using the Line Renderer component, allowing the user to select virtual markers sequentially by pointing the laser at them in numerical order and pressing a button, which fixed each virtual marker in space.

In Step 4, the orientation of the virtual object was calculated based on the first and second fixed virtual markers, and the virtual object was displayed after a button press.



Figure 1. Phantom



Figure 2. Magnetic marker



Figure 3. Virtual object placement

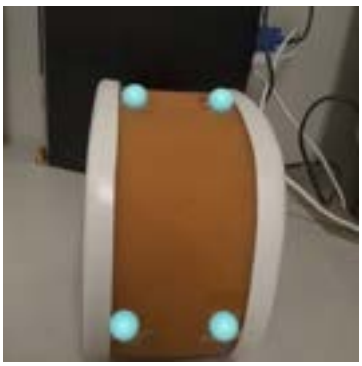


Figure 4. Method① placement



Figure 5. Positional marker



Figure 5. Marker placement order

3 EXPERIMENTAL METHOD

This experiment investigated whether Method ①, in which users manually place markers aligned to reference points, or Method ②, in which positional markers are overlaid on reference points to enable automatic marker placement, allows users to operate more efficiently and achieve higher alignment accuracy.

In addition to the previously developed application—where users visually place markers and placement is accepted by the system within a tolerance of 2 mm—an application that places markers on positional markers was newly developed. Participants operated both applications and evaluated them using a five-point Likert scale across the following three criteria: ease of operation, difficulty of marker placement, and visual naturalness. Additionally, participants rated the degree of alignment between the virtual object and the physical object based on whether the black target in Fig. 3 and the white target of the virtual object visually overlapped. After completing all tasks, participants compared the two applications, and the results were analyzed and validated.

Across all experiments, the placement location of the physical phantom used for overlaying the virtual object was fixed. Operation time was also measured to compare efficiency. Twenty-two participants in their twenties, all familiar with operating information devices, participated in the experiment. All participants received an explanation of the experiment in advance and provided informed consent.

The experimental procedure for both Method ① and Method ② was as follows:

- Explanation of the experiment and acquisition of informed consent

- Wearing the HMD and performing operation practice
- Start of the task upon a timing signal
- Generation of phantom coordinates in virtual space by aligning four reference points and placement of the virtual object
- Completion of evaluations

This study was approved by the Ethics Review Committee.

4 RESULTS AND DISCUSSION

4.1 Positional Error

Data from one participant were excluded due to outliers, and analysis was conducted using data from the remaining 21 participants. In this study, positional error was defined as the difference between the centroid of the four placed marker coordinates and the reference coordinate. The reference coordinate was defined as the average centroid obtained from ten placements performed by an expert, in which visual confirmation verified that the physical and virtual objects were properly aligned, as shown in Fig.3.

The results for coordinate error and operation time are shown in Table 1. One-way analysis of variance revealed no significant difference in the Z-axis error; however, for both the X- and Y-axis errors, Method ② demonstrated significantly smaller errors, with statistically significant differences ($p < 0.05$). Furthermore, the variance of coordinate values was smaller for Method ② across all axes. Operation time was also significantly shorter for Method ② ($p < 0.05$).

These results indicate that placement using positional markers is effective in reducing alignment errors in the X and Y directions.

Table 1. Mean values of coordinate error and operation time and significant difference

	Method①	**Method②**	**p-value**
X directions(mm)	41.1	15.8	0.013*
Y directions(mm)	36.3	3.1	0.001**
Z directions(mm)	30.3	22.2	0.267
Operation time(s)	238.3	72.3	0.000***

*p<.05, **p<.01, ***p<.001

4.2 Usability

For the five-point evaluation conducted after the experiment, the Wilcoxon signed-rank test was applied. For ease of operation, the mean score was 2.8 for Method ① and 4.2 for Method ②, showing a statistically significant difference ($p < 0.05$). In the direct comparison, 77% of

participants rated Method ② as easier to operate, indicating that positional marker–based placement effectively simplifies operation(Table 2).

Table 2. Mean Ratings and *p*-Values for the "Ease of Operation" and "Ease of Marker Placement" Items

	Method①	Method②	p-value
Ease of operation	2.8	4.3	0.002**
Ease of Marker Placement	2.0	4.0	0.000***

*p<.05, **p<.01, ***p<.001

4.3 Visual Discomfort

For the item "Naturalness of the Display," the mean rating was 3.1 for Method ① and 4.0 for Method ②. For the item "Consistency between the Real Object and the Virtual Object," the mean ratings were 3.0 for Method ① and 4.1 for Method ②. A statistically significant difference was observed between the two methods ($p < 0.05$). In the overall comparison, 77% of the participants evaluated Method 2 as producing less visual discomfort than Method ① (Table 3).

However, several participants commented that the virtual object displayed using Method ② appeared to be tilted along the Z-axis compared with Method ①. Since Method ② places the markers one at a time, it provides greater freedom in the object's orientation than Method ①, which places all four markers simultaneously. This increased degree of freedom in orientation is considered to have resulted in the perceived tilt.

Table 3. Mean Ratings and *p*-Values for the " Perceived Discomfort" and "Consistency between the Virtual and Real Objects" Items

	Method①	Method②	p-value
Perceived Discomfort	3.1	4.0	0.005**
Consistency	3.0	4.1	0.000***

*p<.05, **p<.01, ***p<.001

5 CONCLUSION

To address the positional error introduced in the previous workflow—where physical markers were first detected by the application and then aligned—this study proposed and successfully

implemented a method in which positional markers are overlaid on spherical physical markers. The proposed method was effective in reducing alignment error and also simplified operation, resulting in shorter operation times.

As future work, Method ② exhibited tilt in the virtual object because markers were placed one by one. To address this limitation, we plan to develop a function that allows the orientation to be specified and enables accurate display using only a single marker.

Furthermore, improving the accuracy and usability of MR-based visualization systems is expected to reduce visual discomfort and cognitive workload, thereby enhancing users' confidence during medical procedures. Consequently, the proposed method has the potential to contribute to safer and more reliable medical care.

REFERENCES

[1] T. Yoshimitsu, “Utilization of XR Technologies in the Medical Field,” Journal of Medical Instrumentation, vol. 91, no. 5, pp. 457–461, 2021.

[2] H. Oyama, “Current Status and Future Prospects of Medical Applications of XR Technologies,” Journal of the Society of Biomechanisms, vol. 47, no. 3, pp. 145–149, 2023.

[3] T. Nakaguchi, “Trends in XR Technologies Applied to the Medical Field,” IEICE Fundamentals Review, vol. 16, no. 3, pp. 167–175, 2023.

[4] M.Irie, "The Potential of XR: A New World Enabled by High-Fidelity Value Communication," *Knowledge Creation and Integration*, Special Issue: *New Frontiers of Digital Transformation (DX)*, Nomura Research Institute (NRI), Mar. 2021.

EFFECTS OF ECO-LABEL VISUAL MOTIFS ON CONSUMER EVALUATION OF OTC MEDICINE

Jongwon LIM [a] and Shinichi KOYAMA [b*]

[a] *Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan*

[b] *Institute of Art and Design, University of Tsukuba, Japan*

ABSTRACT

Eco-label functions as a visual cue that signals environmental attributes. While eco-label has been widely examined in food, its application in over-the-counter (OTC) medicine remains limited. This study examined how eco-label visual motifs are related to consumer evaluation of OTC medicine. An online survey included 733 participants in Japan. Participants evaluated OTC medicine package images with six label conditions: five eco-label conditions and one no-label condition. The five eco-label conditions consisted of the food carbon footprint of products (CFP) label used in Japan and four visual motif labels using animal, earth, human, and plant motifs. Consumer evaluation was measured using six items: environmentally friendly impression, environmental contribution expectation, environmental impact awareness, purchase intention, recommendation intention, and interest toward OTC medicine. Exploratory factor analysis indicated that the six items formed a single factor; therefore, their mean was used as Perceived Environmental Benefit (PEB). A repeated-measures ANOVA with Greenhouse–Geisser correction showed a significant effect of label condition on PEB, $F(2.02, 1481.86) = 151.05$, $p < .001$, $\eta G^2 = .042$. OTC medicine with all eco-label conditions showed significantly higher PEB than that with the no-label condition. Compared with OTC medicine with the food CFP label, OTC medicine with the animal and human motif labels showed lower PEB, whereas those with the earth and plant motif labels showed comparable PEB. These findings suggest that consumer evaluation of environmental information on OTC medicine packages may differ by visual motif, providing an exploratory basis for eco-label design.

Keywords: *OTC medicine, eco-label, visual motif, carbon footprint, information design*

[*] Corresponding author. 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574, Japan. skoyama@geijutsu.tsukuba.ac.jp

1 INTRODUCTION

Eco-label can function as an informational cue regarding product environmental characteristics and support decision making for environmentally significant products in sustainability communication (Gorton et al., 2021; Thøgersen et al., 2010). Eco-label has been widely examined in food, where environmental information such as carbon footprint of products (CFP) can be presented on packages. In contrast, eco-label design for over-the-counter (OTC) medicine has received limited attention. In Japan, OTC medicine is also purchased directly by consumers, but the use of eco-label on OTC medicine packages is not yet common. This gap suggests that OTC medicine should be examined as a product context for eco-label design.

In addition, visual expression is also important in eco-label design. Prior studies suggest that visual and verbal communication cues and eco-label format can influence consumer responses to environmental information (Rihn et al., 2019; Tang et al., 2004). In addition, presentation format can affect consumers' recall and interpretation of environmental claims, especially when claims are less familiar (Buvár et al., 2026). These points suggest that eco-label visual motifs used for OTC medicine may be related to consumer evaluation.

Accordingly, this study examined how eco-label visual motifs are related to consumer evaluation of OTC medicine by addressing two research questions: RQ1. Does OTC medicine with eco-label conditions receive higher consumer evaluation than OTC medicine with the no-label condition? RQ2. How do visual motif labels for OTC medicine differ in evaluation compared with the food CFP label?

2 METHODS

An online survey was conducted with 733 participants in Japan. Eligible participants were aged 30–49 years, lived in Japan, and had experience purchasing OTC medicine. The final sample included 481 men and 252 women (mean age = 41.94, SD = 5.13). We asked participants to evaluate OTC medicine package images with six conditions: five eco-label conditions and one no-label condition. The five eco-label conditions consisted of the food CFP label condition and four visual motif label conditions with animal, earth, human, and plant motifs. The food CFP label currently used in Japan was included as a comparison because no commonly used CFP label for OTC medicine was available. Three prototypes were prepared for each visual motif, resulting in 12 prototypes. These conditions differed only in the visual motif, while the label frame, color, font, font size, and text were kept constant. The presentation order of the images was randomized.

We measured consumer evaluation using six items: environmentally friendly impression, environmental contribution expectation, environmental impact awareness, purchase intention, recommendation intention, and interest toward OTC medicine (1 = strongly disagree; 5 = strongly agree). Exploratory factor analysis indicated that the six items formed a single factor, and internal consistency was high (Cronbach's α = .925). Therefore, their mean was used as Perceived Environmental Benefit (PEB). A repeated-measures ANOVA was conducted with label condition as the within-subject factor, with Greenhouse–Geisser correction applied for sphericity violation. Holm-adjusted pairwise comparisons were then conducted.

3 RESULTS

3.1 Comparison of OTC medicine with eco-label conditions and the no-label condition

For RQ1, we examined whether OTC medicine with eco-label conditions received higher consumer evaluation than that with the no-label condition. A repeated-measures ANOVA with Greenhouse–Geisser correction showed a significant effect of label condition on PEB, $F(2.02, 1481.86) = 151.05$, $p < .001$, $\eta G^2 = .042$. Holm-adjusted pairwise comparisons indicated that each eco-label condition had significantly higher PEB than the no-label condition ($p < .001$). The mean difference from the no-label condition (M = 2.73, SD = 0.74) ranged from 0.357 for the human motif label condition to 0.484 for the earth motif label condition.

3.2 Comparison of OTC medicine within eco-label conditions

For RQ2, OTC medicine with the food CFP label was used as a comparison point for examining the evaluation patterns of those with the visual motif label conditions (Figure 1). OTC medicine with the food CFP label received a mean PEB of 3.20 (SD = 0.81). In relation to this condition, OTC medicine with the animal motif label (M = 3.12, SD = 0.82) and that with the human motif label (M = 3.09, SD = 0.82) showed lower PEB, whereas that with the earth motif label (M = 3.21, SD = 0.82) and that with the plant motif label (M = 3.18, SD = 0.80) showed comparable PEB. Holm-adjusted comparisons showed lower PEB for the animal and human motif label conditions ($p < .001$), but no significant differences for the earth and plant motif label conditions.

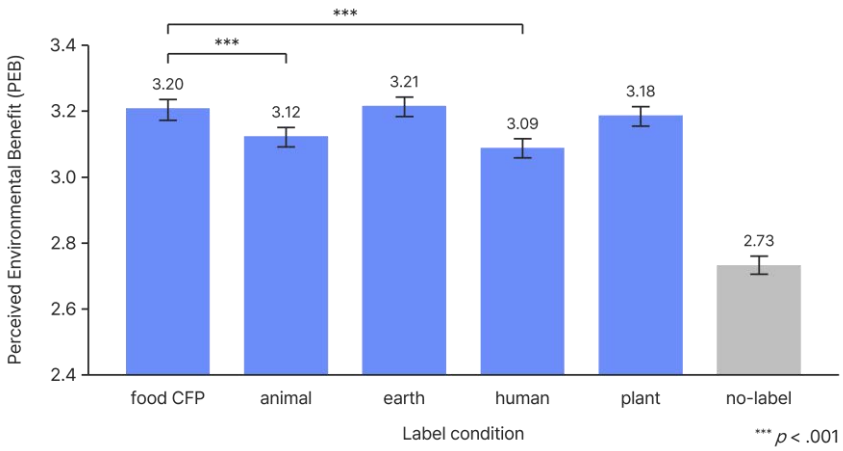


Figure 1. Comparison of OTC medicine within eco-label conditions using a repeated-measures ANOVA

4 DISCUSSION AND CONCLUSION

For RQ1, the result suggests that OTC medicine with eco-label conditions can be associated with more positive consumer evaluation, as represented by PEB. This finding can be discussed alongside prior research describing eco-label as an informational cue regarding product environmental characteristics and as a tool for supporting consumer decision making for environmentally significant products (Gorton et al., 2021; Thøgersen et al., 2010). Thus, this result suggests eco-label design for OTC medicine is worth further examination as a product context.

For RQ2, OTC medicine with the food CFP label provided a useful comparison point for examining visual motif label conditions. OTC medicine with the earth and plant motif labels showed comparable PEB, whereas those with the animal and human motif labels showed lower PEB. This pattern suggests that visual motif labels cannot be assumed to improve consumer evaluation simply because they are visually adapted to OTC medicine. Rather, consumer evaluation may differ depending on the visual motif used. This result can be discussed in relation to prior studies showing that visual and verbal cues and eco-label format can influence consumer responses to environmental information (Rihn et al., 2019; Tang et al., 2004). It also relates indirectly to research showing that presentation formats can affect consumers' recall and interpretation of less familiar environmental claims (Buvár et al., 2026).

Taken together, these findings suggest that eco-label design for OTC medicine is worth further examination in sustainability communication and that consumer evaluation of environmental information on OTC medicine packages may differ depending on the visual motif used. This study provides an exploratory basis for examining visual motifs in OTC medicine eco-label design.

ACKNOWLEDGMENTS

This study was supported by JST SPRING (Grant Number JPMJSP2124) to Jongwon Lim, and JSPS Grant-in-Aid for Scientific Research (C) (Grant Number 25K15773) to Shinichi Koyama.

REFERENCES

Buvár, Á., Berezvai, Z., & Dúll, A. (2026). Decoding sustainability: The effect of eco-labels and textual cues on consumers' spontaneous recall and interpretation of environmental claims. *Sustainable Production and Consumption*, *63*, 203–220. https://doi.org/10.1016/j.spc.2026.01.008

Gorton, M., Tocco, B., Yeh, C. H., & Hartmann, M. (2021). What determines consumers' use of eco-labels? Taking a close look at label trust. *Ecological Economics*, *189*, 107173. https://doi.org/10.1016/j.ecolecon.2021.107173

Rihn, A., Wei, X., & Khachatryan, H. (2019). Text vs. logo: Does eco-label format influence consumers' visual attention and willingness-to-pay for fruit plants? An experimental auction approach. *Journal of Behavioral and Experimental Economics*, *82*, 101452. https://doi.org/10.1016/j.socec.2019.101452

Tang, E., Fryxell, G. E., & Chow, C. S. F. (2004). Visual and verbal communication in the design of eco-label for green consumer products. *Journal of International Consumer Marketing*, *16*(4), 85–105. https://doi.org/10.1300/J046v16n04_05

Thøgersen, J., Haugaard, P., & Olesen, A. (2010). Consumer responses to ecolabels. European *Journal of Marketing*, *44*(11-12), 1787–1810. https://doi.org/10.1108/03090561011079882

EFFECTS OF VISUAL SELF-REFERENTIAL FEEDBACK ON BODY OWNERSHIP AND USABILITY IN HMD-BASED ROBOT TELEOPERATION

Dokyung KIM [a], and Shinichi KOYAMA [b*]
[a] *Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan*

[b] *Institutue of Art and Design, University of Tsukuba, Japan*

ABSTRACT

This study examines how visual self-referential conditions influence body ownership and affective responses during remote robot operation using a head-mounted display (HMD). In the experiment, 24 participants (aged 21–34) were recruited. Participants placed their chin on a chin-rest, wearing an HMD connected to a camera mounted on a robot, which transmitted real-time forward-view images. They moved the robot (a modified radio-controlled toy car) in a straight line within 3 seconds toward a target using control levers. Participants were divided into two groups, a self-referential group and an external-object group. Both groups completed three robot-operation trials using self-face feedback and two different types of dolls: a lion doll (A) and a penguin doll (B). In the self-referential group, participants first moved the robot toward doll A, second toward their own face wearing HMD, and third to doll A again. In the external-object group, participants first moved the robot toward doll A, second toward a doll B, and third to doll A again. After the first and third run, both groups completed a questionnaire assessing body ownership (self, other, separate) and usability (ease, difficulty, control, comfort, stress, enjoyment, boredom). The differences in scores between the first and last questionnaire was computed and independent-samples t-tests were conducted to examine group differences. Participants in self-referential group who directed the robot toward their own face showed a significant decrease in body ownership ($p < .05$), along with a tendency toward increased stress and decreased enjoyment. These findings suggest that presenting one's own face may reduce body ownership. Body ownership may reduce stress and increase enjoyment. The results are consistent with prior studies showing that body ownership arises from multisensory integration (Botvinick & Cohen, 1998; Ehrsson, 2007). Robot interface design adjusting the sense of body ownership through visual self-referential conditions may change user experience.

Keywords: *Robot Teleoperation, Body Ownership, Usability, Visual Self-Referential Condition*

* Corresponding author. 1-1-1 Tennnodai, Tsukuba, Ibaraki, Japan. skoyama@geijutsu.tsukuba.ac.jp

1 INTRODUCTION

In human–robot interaction (HRI), improving user experience has become a central research objective. With advances in virtual reality (VR) and head-mounted display (HMD) technologies, robots are no longer used merely as remote control devices. In particular, body ownership—perceiving a robot as part of one's own body—is gaining attention. Previous research suggests that embodied experiences are closely related to body ownership and the sense of agency, including feelings of control over bodily actions (Kalckert & Ehrsson, 2012). Previous cognitive neuroscience research suggests that body ownership can extend to teleoperation contexts through multi-sensory integration (Lenggenhager et al., 2007). For example, the Rubber Hand Illusion showed that synchronized visual and tactile stimulation can induce body ownership toward a rubber hand (Botvinick & Cohen, 1998; Ehrsson, 2007). In HRI research, human-like virtual bodies have been shown to enhance body ownership (Shin, 2020). However, prior research has not clearly compared self-body and external-object experiences in robot teleoperation. This study examines whether an intermediate stimulus under self-referential and external-object conditions influences users' body ownership and usability responses in HMD-based robot teleoperation. Participants were divided into two groups and completed three trials using self-face feedback and two different types of dolls as stumuli: a lion doll (A) and a penguin doll (B). Self-referential group moved the robot toward doll A, then toward the participant's own face viewed through the HMD, and doll A again. External-object group moved the robot toward doll A, doll B, and doll A again. After Trials 1 and 3, both groups completed a questionnaire on body ownership, and usability. Change scores (Trial 3 minus Trial 1) were compared between groups to examine which group showed greater changes. An independent-samples t-test was conducted to examine whether these differences were statistically significant.

2 METHOD

2.1 Participants

A total of 24 university students were recruited through on-campus advertisements at the University of Tsukuba. All participants completed the experiment (age range: 21–34 years; 17 males, 7 females). Only one participant had expert-level knowledge of robotics, while the others reported limited or no prior experience.

2.2 Environment

Participants received an information sheet describing the study purpose, procedures, duration, risks, data handling, and voluntary nature of participation. They were informed of their right to withdraw at any time without penalty. All participants reported having normal or corrected-to-normal vision. A chair and a chin rest were used to stabilize the participants' heads. A straight cardboard corridor (length: 120 cm) was placed in front of each participant for the robot to move through. The robot was a modified radio-controlled car (6 × 10 cm) equipped with a camera. Its outer shell was 3D-printed and covered with a fur-like textile. The camera was connected to an HMD (Big Eyes 2), providing real-time forward visual feedback. Part of the fur was visible in the camera view, reinforcing the perception of the robot as a non-human external object (Figure 1).



Figure 1. A modified radio-controlled car equipped with a camera

2.3 Procedures

Participants sat and placed their chin on a chin rest. They wore an HMD that provided real-time forward visual feedback from a camera mounted on the robot. They were instructed to use a controller to move the robot forward and backward. Participants moved the robot in a straight line through the corridor toward a stimulus within 3 secsonds with a 40 BPM metronome. Three stimuli were used: a lion doll (A; 70 × 100 mm), a penguin doll (B; 120 × 70 mm), and the participant's own face viewed via the HMD. Participants (n = 24) were divided into two groups: Self-referential group (n = 12) and External-object group (n = 12). Both groups completed three trials and questionnaires on body ownership and usability after Trials 1 and 3. In the self-referential group, participants moved the robot toward doll A (Trial 1), their own face viewed through the HMD (Trial 2-1), and doll A again (Trial 3). In the external-object group, participants moved the robot toward doll A (Trial 1), doll B (Trial 2-2), and doll A again (Trial 3). For both groups, change scores between Trials 1 and 3 were calculated (Figure 2, Figure 3).

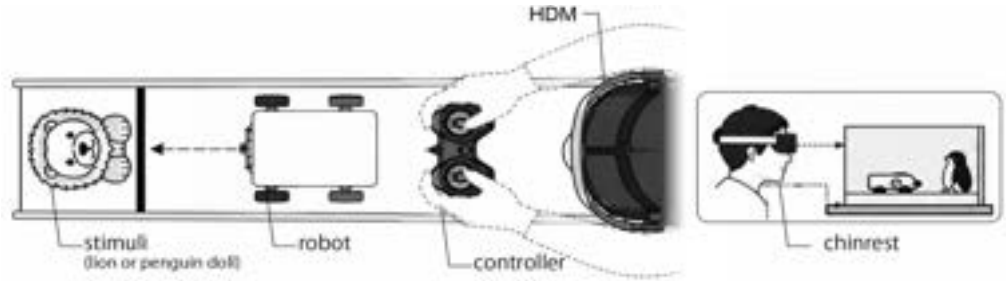


Figure 2. Overview of the HMD-Based Robot Teleoperation Experiment

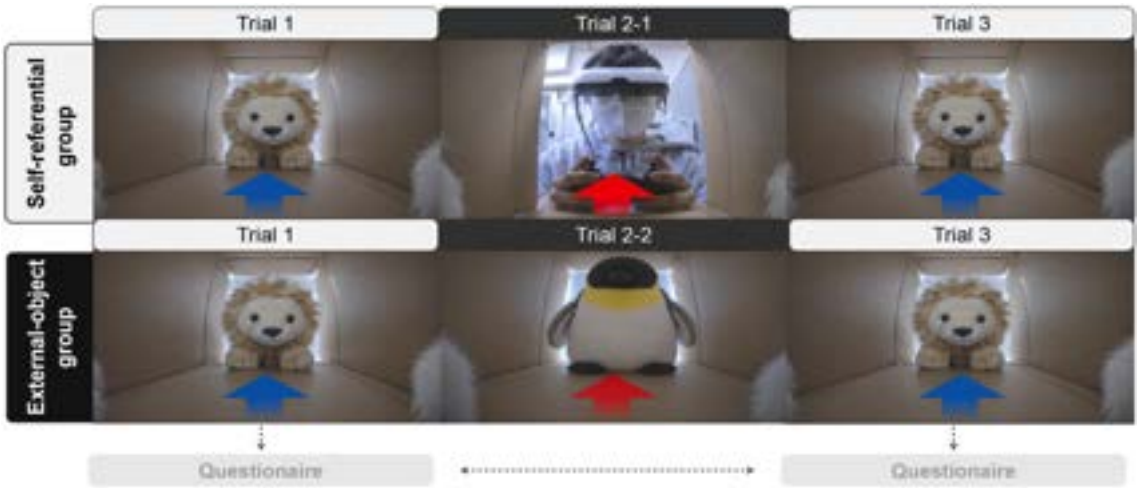


Figure 3. Three trials of robot operation in self-referiantial group and external-object group

2.4 Questionaire

Both groups completed the same questionnaire after Trials 1 and 3. The questionnaire consisted of 10 items assessing body ownership (self, other, separate) and usability (ease, difficulty, control, comfort, stress, enjoyment, boredom). Items included the following questions: (Q1) To what extent did you feel that the robot was part of your own body? (Q2) To what extent did you feel that the robot was another person's body? (Q3) To what extent did you feel that the robot was a separate entity from yourself? (Q4) How easy was it to operate the robot? (Q5) How difficult or complex was it to operate the robot? (Q6) To what extent did you feel in control of the robot? (Q7) How comfortable did you feel during the operation? (Q8) How much stress did you experience during the operation? (Q9) How enjoyable was the operation? (Q10) How bored did you feel during the operation? Responses were recorded on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The experiment lasted approximately 20–30 minutes, after which participants reported demographic information and prior experience with robotics. Items were adapted from established embodiment scales and prior HRI studies (Longo et al., 2008; Kalckert & Ehrsson, 2012; Bartneck et al., 2009).

2.5 Analysis

The data were analyzed using IBM SPSS Statistics 29.0. For each participant, change scores were computed for each item as Trial 3 minus Trial 1. These change scores were compared between two Groups to examine which group showed greater changes. An independent-samples t-tests were conducted to assess whether between-group differences were statistically significant.

3 RESULTS

All participants completed the same 10-item questionnaire on body ownership and usability after Trials 1 and 3. Change scores (Trial 3 − Trial 1) were compared between the self-referential group and the external-object group using independent-samples t-tests (Table 1). Embodiment (Q1) showed a significant decrease in the self-referential group (ΔM = −0.91, t = −2.27, p = .03). Other-body sensation (Q2) increased, but was not significant (ΔM = 0.50, t = 1.01, p = .32). Perceived separateness (Q3) increased, but the change was not significant (ΔM = −0.09, t = −0.17, p = .87). Ease of operation (Q4) decreased, but the change was not significant (ΔM = −0.16, t = −0.48, p = .64). Perceived complexity (Q5) decreased, but the change was not significant (ΔM = −0.25, t = −0.52, p = .61). Sense of control (Q6) increased, but the change was not significant (ΔM = −0.09, t = −0.17, p = .87). Comfort (Q7) increased, but the change was not significant (ΔM = 0.00, t = 0.00, p = 1.00). Stress (Q8) increased, but the change was not significant (ΔM = 0.17, t = 0.48, p = .63). Enjoyment (Q9) decreased, but the change was not significant (ΔM = −0.25, t = −0.81, p = .43). Boredom (Q10) increased, but the change was not significant (ΔM = 0.00, t = 0.00, p = 1.00). Only embodiment (Q1) showed a significant decrease (p < .05). All other measures showed no statistically significant changes (all ps > .05). Other-body sensation and perceived separateness increased, whereas ease of operation and perceived complexity decreased. Sense of control, comfort, stress, and boredom increase, while enjoyment decreased.

Table 1. Independent-samples t-test results (Trial 3 -Trial 1)

variable		Self-referential(M, SD)		External-object(M, SD)		ΔM	$t(22)$	p
		pre	post	pre	post			
body ownership	embodiment (Q1)	3.25, 0.97	2.67, 1.23	2.92, 0.88	3.25, 0.97	-0.91	-2.27	.03*
	other-body (Q2)	2.75, 0.97	3.58, 1.00	2.67, 1.03	3.00, 0.74	0.50	1.01	.32
usability	separate (Q3)	3.33, 1.15	3.92, 1.08	3.50, 0.97	4.17, 0.57	-0.09	-0.17	.87
	ease (Q4)	3.58, 1.31	3.50, 1.17	3.17, 1.03	3.25, 1.07	-0.16	-0.48	.64
	complexity (Q5)	2.25, 0.97	2.08, 1.08	2.08, 0.88	2.17, 0.94	-0.25	-0.52	.61
	control (Q6)	3.75, 1.14	3.83, 0.94	3.67, 0.95	3.83, 0.67	-0.09	-0.17	.87
	comfort (Q7)	2.92, 1.24	3.17, 1.11	2.92, 0.99	3.17, 0.92	0.00	0.00	1.0
	stress (Q8)	1.83, 0.58	2.50, 1.00	2.08, 0.74	2.58, 0.97	0.17	0.48	.63
	enjoyment (Q9)	4.00, 0.95	3.92, 1.00	3.42, 0.85	3.58, 0.70	-0.25	-0.81	.43
	boredom (Q10)	1.75, 0.62	2.00, 0.85	2.00, 0.94	2.25, 0.95	0.00	0.00	1.0

4 DISCUSSION

This study examined how visual self-referential conditions influence body ownership and usability in HMD-based robot teleoperation. A total of 24 participants were divided into two groups and completed three trials under either a self-face condition or an external-object condition. In the self-referential group, the robot was moved toward doll A, the participant's own face viewed through the HMD, and doll A again. In the external-object group, the robot was moved toward doll A, doll B, and doll A again. After Trials 1 and 3, participants completed a questionnaire assessing body ownership (self, other, separate) and usability (ease, difficulty, control, comfort, stress, enjoyment, boredom). Change scores (Trial 3 minus Trial 1) were compared between groups to examine which condition produced greater changes. Independent-samples t-tests were conducted to examine whether these differences were statistically significant. Previous HRI research has suggested that embodiment experiences during robot interaction may influence users' body ownership and usability, and that such mechanisms may extend to teleoperation contexts (Lenggenhager et al., 2007). Previous research has also reported that human-like virtual bodies can enhance body ownership (Shin, 2020). The present findings suggest that the self-referential condition may have reduced embodiment by shifting the robot perception as a human-like virtual body from a self-body to an external object. Increases in stress and boredom, along with decreased enjoyment, may reflect intensified self-awareness and emotional tension under the self-face condition.

5 LIMITATIONS

This study is limited by its small sample size, and the observed trends should be interpreted as preliminary. The experiment used a simple linear movement task and static stimuli, limiting generalizability to more complex interactive contexts, such as obstacle avoidance, directional changes, navigation through dynamic environments, and the execution of diverse tasks. Possible learning or fatigue effects across repeated trials may not have been fully controlled. The novelty effect of viewing one's own face as an experimental stimulus may simply reflect random variation.

6 CONCLUSION

In HMD-based robot teleoperation, embodiment significantly decreased only when participants viewed their own faces rather than external objects, suggesting that self-referential experiences may reduce robot embodiment and body ownership. Increased stress and boredom suggest that self-referential conditions may regulate user–robot affective distance. Depending on visual feedback conditions, the user–robot relationship can dynamically shift between perceiving the robot as an extension of the one's own body and as a separate external object. Users' affective experiences and psychological engagement during robot teleoperation may also be influenced.

ACKNOWLEDGMENTS

This study was supported by JSPS Grant-in-Aid for Challenging Research (Exploratory) (Grant Number 20K21823) to Shinichi Koyama.

REFERENCES

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. Nature, 391(6669), 756. https://doi.org/10.1038/35784

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. Science, 317(5841), 1048. https://doi.org/10.1126/science.1142175

Kalckert, A., & Ehrsson, H. H. (2012). Moving a rubber hand that feels like your own: A dissociation of ownership and agency. Frontiers in Human Neuroscience, 6, Article 40. https://doi.org/10.3389/fnhum.2012.00040

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. Science, 317(5841), 1096–1099. https://doi.org/10.1126/science.1143439

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. Cognition, 107(3), 978–998. https://doi.org/10.1016/j.cognition.2007.12.004

Preston, C., & Ehrsson, H. H. (2014). Illusory changes in body size modulate body satisfaction in a way that is related to non-clinical eating disorder psychopathology. PLOS ONE, 9(1), e85773. https://doi.org/10.1371/journal.pone.0085773

Preston, C., Kuper-Smith, B. J., & Ehrsson, H. H. (2015). Owning the body in the mirror: The effect of visual perspective and mirror view on the full-body illusion. Scientific Reports, 5, Article 18345. https://doi.org/10.1038/srep18345

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators and Virtual Environments, 10(3), 266–281. https://doi.org/10.1162/105474601300343603

EMOTION-SCAFFOLDED DESIGN EDUCATION AND KANSEI LEARNING ENVIRONMENTS

AMIC G. Ho, HONG KONG METROPOLITAN UNIVERSITY, HONG KONG

ABSTRACT

Upon the framework of Emotion and Design, the concept of Kansei learning environments is a deliberate construction of emotional ecologies designed to facilitate the affective trajectories of students throughout the design process. The paper addresses long-standing issues for junior design students who often experience anxiety, frustration, and disengagement from open-ended assignments where unmanaged emotions inhibit exploration, risk-taking, and reflective judgement. A mixed-methods intervention was developed and implemented within two undergraduate communication design studios (n ≈ 60), which incorporated emotionally scaffolded pedagogy through the use of structured emotional check-ins, Kansei-based reflection prompts, visualisations of emotional momentum, and redesigned critique processes that emphasised developing emotional literacy in addition to providing formal and conceptual feedback on student projects. Data collection methods included quantitative self-report measures of students' emotional state, perceived self-efficacy, and studio engagement in conjunction with rubric-based assessments of creativity, coherence, and craftsmanship in the final student products. Additionally, qualitative data was collected using emotion diaries, observational research in the studio environment, and post-project interviews with students regarding their understanding and utilisation of emotion as a design resource. Early analysis indicates that the emotion-scaffolded studios resulted in increased emotional granularity, increased persistence in times of difficulty, and more intentional usage of affective cues in the making of design decisions related to improved assessments of creative performance.

This research study adds to prior literature by (1) offering an empirical definition of Kansei learning environments based on the 3E model; (2) providing empirical design principles for incorporating emotion scaffolding into studio curriculum design; and (3) addressing how emotion-aware pedagogy can broaden the scope of Kansei engineering from product experience to the performance of learning, teaching and creative practice in design school contexts.

Keywords: *Kansei learning environments, emotion-scaffolded pedagogy, design education, emotional granularity, creative self-efficacy*

1 INTRODUCTION

The inclusion of emotion in design practice and the user experience have been recognised since a long time ago. Yet the development of the inclusion of emotion systematically into design

education remains underdeveloped. In particular, it relates to the emotional trajectories students undergo during their project lifecycles. Many junior design students express frustration, anxiety, and disengagement when they encounter open-ended assignments, short deadlines and vague expectations. Such uncontrolled emotions can limit exploration risks, limit conceptual thinking, and impede critical evaluation, thus limiting the quality of the work students produce. This study builds upon the Emotion and Design Framework, and the 3E Model of Emotion Design (Emotional Design) and Emotionalisation of Design. Thus, we construct Kansei learning environment(s) as intentional emotional ecology designed to support students not just by acknowledging their feelings but also by using these as a basis for their own thinking, making, and evaluating in a studio context. Based upon this view of emotion, we consider emotion as a medium and as a motivator for learning. Emotion affects students' focus, motivation, persistence and capacity to make subtle judgments concerning aesthetics and concepts.

In response to this need, our research provides a mixed-methodology-based pedagogical intervention in two undergraduate communication design studios (n ≈ 60). We incorporated emotion-scaffolding methods into the routine activities of each studio. These methods include scheduled emotional check-ins, Kansei-based reflection questions, diagrams that visualise 'emotional momentum' throughout a student's project timeframe, and revised critique processes where emotional intelligence be evaluated equally as formally and conceptually. We assess quantitative data regarding students' emotional states, perceptions of self-efficacy, and degree of engagement in studio activities, along with rubrics used to evaluate creativity, coherence, and craftsmanship. Qualitative information from students' emotion diary entries, studio observation notes, and post-project interview responses help us understand how students perceive and utilise emotion as a resource for designing. Our ultimate goal is to develop evidence-based definitions for Kansei learning environments developed through the 3E Model and create design principles for implementing emotion-scaffolding in design studio curriculum, and to argue that Kansei Engineering should evolve beyond the scope of users' experiences of products toward the experiences of students performing learning, teaching, and creative practice within design schools.

2 EMOTION-SCAFFOLDED DESIGN EDUCATION AND KANSEI LEARNING ENVIRONMENTS

2.1 The Theoretical Foundations of Emotion and Design

Ho and Siu's (2012) three dimensional model of emotion and design: emotion design, where emotional issues are used by the designer to create interaction with the user; emotional design, where artefacts elicit an emotional response from users; emotionalise design, where designers' emotional influences lead to the way they carry out their work. These three dimensions provide a holistic view of how emotions impact all stages of the design process. From idea generation and creation to post-design evaluation and use. In addition, Ho (2014) provided empirical support for the fact that designers' emotional experience can improve designers' ability to gather information and make decisions regarding the design process. Thus, Ho (2014) challenged previous notions that emotional experience was a barrier to rational thought processes.

2.2 Kansei Engineering and Its Application to Learning Environments

Kansei engineering is a Japanese engineering approach to understand and measure consumers' emotional responses towards various aspects of products. In recent years, there has been some interest in using Kansei engineering in educational contexts. However, this application remains largely confined to computer-mediated learning technologies. Chuah et al. (2008) proposed applying Kansei methodology to a virtual reality-based learning environment. They argued that because of its ability to evoke emotions in the learner, it requires systematic consideration via frameworks that relate learners' emotional responses to the specific design features of the learning environment. So far, however, these approaches remain largely underdeveloped in face-to-face design studio education settings. Design studio education often occurs over many weeks, involving multiple iterations of the design process, critique sessions, peer group collaborative activities, and hands-on exploration of materials (Boling, Gray & Smith, 2020).

2.3 Creative Self-Efficacy and Its Influence on Design Performance

A growing body of literature has identified that 'emotional scaffolding' or designing instruction with consideration for students' emotions, is an important educational strategy. As such, affective pedagogy acknowledges that emotions are produced from a variety of physiological, cognitive, and environmental stimuli. Students' beliefs about their ability to produce novel and useful solutions (i.e., their creative self-efficacy) predict their creative output, persistence in engaging in creative activities, and willingness to participate in divergent thinking (Bandura, 1997). While previous studies using design thinking interventions have shown some effect on creative self-efficacy, they suggest that the lack of focus on the affective aspects of the learning experience contributed to these limited findings. Emotionally scaffolding students' emotional response to ambiguity, failures, and criticism could help decrease how often those experiences result in feelings of ineptitude (Choi, 2004).

2.4 Design Studio Pedagogy: Critique Practices, Cultural Norms, and Student Well-Being

Research has also documented significant levels of anxiety, stress, and mental illness among design students. The authors attribute this to system-related issues (e.g., extremely heavy workload; unclear evaluation rubrics; high-stakes, public design reviews; the culture of working excessively). Critics of traditional design studios argue that these formats, particularly the confrontational format of the design review or jury, continue to promote exclusionary studio cultures based upon tradition and authority rather than evidence-based education. Advocates for redesigning traditional design studio education recommend reframing studio-based education so it supports the development of whole-person students (Mathisen & Brønnick, 2009). Specifically, advocates encourage educators to view feedback/critique as opportunities for dialogue and growth rather than as evaluative tools meant to inspire fear.

2.5 Synthesis: Towards Emotion-Scaffolded Design Education

The current literature provides an array of common findings that help to develop our theory and methods of providing emotion-scaffolds in Design Education within the context of Kansei Learning Environments. Firstly, emotions are not merely supplemental to design thinking, but

are integral to it. Emotions can either positively or negatively impact a person's ability to focus their attention, be motivated, process information, make decisions, and produce high-quality creative results (Saarni, 2017). Secondly, while many students experience both positive and negative emotional responses during the educational design process, little formal assistance has been given to students to support them in developing their emotional intelligence. Thirdly, there are now multiple evidence-based methodologies available for creating emotion-aware pedagogies for supporting students' emotional trajectories as they learn. These include Affective Pedagogy, Emotional Scaffolding, Creative Self-Efficacy Research, and Kansei Engineering. Finally, design studios traditionally promote a culture of extreme hardship and stress where individual perseverance is expected at any cost; this needs to be transformed so that design studios better reflect the current understanding of the relationships among learning, creativity, and well-being (Gray, 2013) .

While research has made great strides toward improving the way we understand the role of emotion in the design process (Ho & Siu, 2012, 2014) and how Kansei Engineering may be applied to create emotionally aware learning environments (Chuah et al., 2008), few studies exist that empirically examine emotion-scaffolding in design education contexts. While Park et al. (2019) provide excellent frameworks for emotionally scaffolding young children's learning, and while Hammond and Moore (2018) establish a general framework for implementing affective pedagogy in higher education, very little research exists concerning how emotional scaffolding functions in the unique cultural environment of the design studio. Therefore, this study proposes to investigate two major areas of importance in addition to developing and evaluating an emotion-scaffolded pedagogical strategy within undergraduate Communication Students'Design Studios. Specifically this study explored the long-term relationship between Creative Self Efficacy and Studio-Based Learning Outcomes and the Longitudinal Impact of Emotion Aware Pedagogies on Students Professional Development and Career Sustainability.

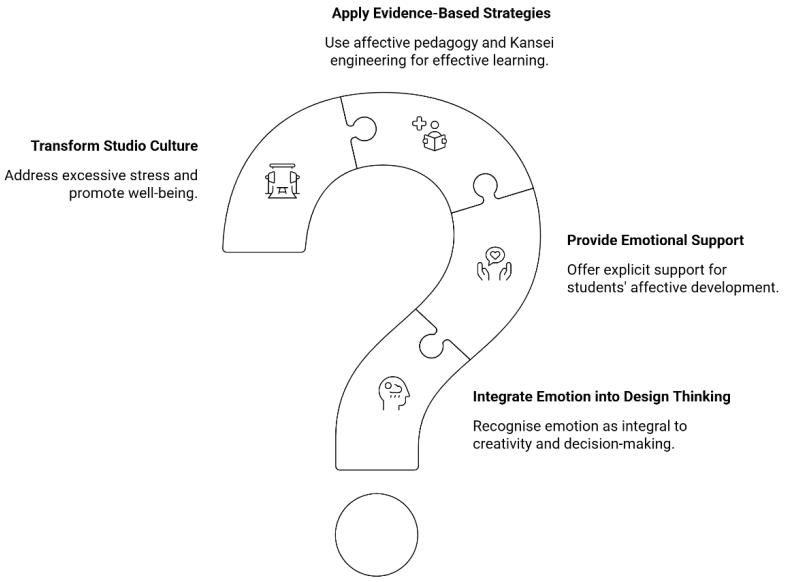


Figure 1. The figure shows the way to implement emotion-scaffolded design education

3 RESEARCH METHODS

This research employed a convergent mixed methods approach to determine if emotion-supported Kansei learning environments positively affect the emotional progression, self-confidence, and creativity of design students during undergraduate communication design studies. Quantitative and qualitative approaches were equally weighted and merged throughout all aspects of the methodology (design/interpretation), allowing for triangulation among participant scales; performance measurements, participants' emotion journals; observations, and participant interviews. Two, second/third year communication design studios at a public institution were used to recruit approximately 60 students (n ≈ 30 per class). To participate in this study, students had to be formally registered for that semester's communication design course, complete at least one other design foundation course, and be able to reflect on their emotional experience while learning. In-class announcements and subsequent emails were used for recruitment purposes. Institutional ethics protocols were followed to obtain participants' written consent. Participants were free to decline participation without any negative consequences on their academic standing.

The intervention was embedded in a regular 12-week semester. Three distinct yet overlapping phases defined the intervention: Baseline Mapping (Weeks 1–2); Emotion-Supported Studio Practice (Weeks 3–10); Consolidation and Reflection (Weeks 11–12). Baseline data regarding students' emotional states, levels of self-confidence, and typical ways in which they engaged in the studio were collected using standardised participant report scales derived from established psychological instruments during Weeks 1 & 2. Students were also required to submit an initial portfolio piece, evaluated using an analytic rubric assessing creativity, conceptual coherence, and craftsmanship, creating a baseline measure of student performance (Jobst & Meinel, 2012). During Weeks 3–10, the emotion-scaffolded Kansei studio became a normal component of the teaching routine. Each studio session included a short emotional check-in. Using a structured rating sheet, participants identified and labeled their emotional state, what they expected to be challenging, and what coping mechanisms would be utilised. After the emotional check-in, participants went through Kansei-based reflection exercises at critical junctures in their projects (research, ideation, prototype development/refinement). These reflection exercises involved having students map how they believed specific design elements (color, typeface, image, composition) would elicit or modify specific emotional responses in themselves and potential users. Critique formats were modified to emphasise emotional intelligence. For example, peer and tutor feedback was no longer limited to comments on formal qualities and conceptual thinking but now incorporated comments on perceived emotional tone and the designers' reported emotional process.

Quantitative data collection was concurrent with studio activities. Self-report measures of emotional state, self-confidence, and studio engagement were administered at four separate times: baseline; mid-project, immediate pre-final critique, and post-project completion. The final project submissions were independently assessed by two raters blinded to the focus of the intervention. This was done to minimise expectation bias. Reliability estimates were computed for the rubric scores before calculating average ratings.

Three types of qualitative data existed for this study: emotion diaries; nonparticipant observation of selected studio sessions; and semi-structured interviews with participants. Students completed weekly emotion journals identifying significant high/low emotional episodes experienced during the semester including apparent causes and strategies they employed to cope with those feelings. A sample of observed studio sessions focused on observable emotional expressions; utilisation of emotion-scaffolding materials; and the evolution of critique interactions. By the end of the semester a purposeful subset of students who represented varied emotional progressions participated in individual interviews discussing their experiences with the Kansei activities and how emotions affected their decision-making processes (Kashdan, Barrett, & McKnight, 2018).

A multi-phase integrative strategy was applied to analyse data. Descriptive statistical techniques and repeated measures analyses were used to evaluate the impact of the intervention on students' emotional states, self-efficacy, engagement and rubric score averages over the duration of the semester. Thematic coding was used to categorise qualitative data from diaries, observations and interviews into thematic categories using deductively-derived codes (i.e., emotional detail; emotional movement; regulatory strategies) as well as inductively-developed themes that emerged from the corpus. Integration occurred through side-by-side displays of quantitative trends juxtaposed against illustrative excerpts from qualitative data; these displays provided a comprehensive narrative explaining how emotion-scaffolding Kansei-based interventions influenced students' creative outcomes and learning experiences.

4 RESULTS

The number of different emotional states that were reported at both the beginning and end of this study rose dramatically. Initially, the average student could report just under 4 types of emotional state ($M = 4.10$, $SD = 1.06$). By the time they had finished the course, however, this had risen to approximately 5 ($M = 4.87$, $SD = 1.14$) according to a paired sample t-test ($t(99) = -12.17$, $p < .001$) (figure 2). The level of anxiety experienced by students also dropped dramatically over the period of this study. On average, students reported experiencing anxiety at a level of 5.42 ($SD = 1.05$) when they began the course. By the time they completed it, they reported experiencing anxiety at a level of around 4.23 ($SD = 1.26$) ($t(99) = 16.36$, $p < .001$) (figure 3). Frustration experienced by students similarly reduced throughout this study. Initially, students averaged 5.04 ($SD = 1.09$) frustration. By completion of the course, this had fallen to around 4.41 ($SD = 1.22$) ($t(99) = 9.86$, $p < .001$)(figure 4).

Curiosity increased as did enjoyment. Initially, students averaged 4.81 ($SD = 0.92$) curiosity. This had increased to 5.73 ($SD = 0.96$) by the end of the course ($t(99) = -15.09$, $p < .001$). Initial enjoyment of the course was measured at an average of 4.53 ($SD = 1.04$). By completion of the course, this had increased to an average of 5.81 ($SD = 0.93$) ($t(99) = -18.35$, $p < .001$). In terms of how well students were able to identify different emotional states, we refer to these as emotional granularity scores. These scores increased significantly over the duration of this study. When students started the course, they could, on average, identify slightly less than 4 different emotional states ($M = 3.93$, $SD = 1.25$). By completion of the course, this figure had risen to

around 6.64 (SD = 1.73) ($t(99) = -24.86$, $p < .001$) (figure 5). In other words students ability to describe different emotional states improved over the duration of this study by approximately two and a half additional emotional states. During critique sessions where students provided feedback on each other's designs, there was considerable use of emotion-related vocabulary (average score for use of emotion vocabulary is M = 5.58, SD = 0.99) (figure 6).

Student perceptions regarding their own abilities to complete design tasks increased considerably over the duration of this study. Initially, students believed they were able to complete design tasks with an effectiveness of around 4.23 out of five possible points (SD = 0.79). By the end of the course, students indicated they believed they would be successful in completing similar design tasks with an effectiveness rating of around 5.28 out of five possible points (SD = 0.95) ($t(99) = -17.95$, $p < .001$).

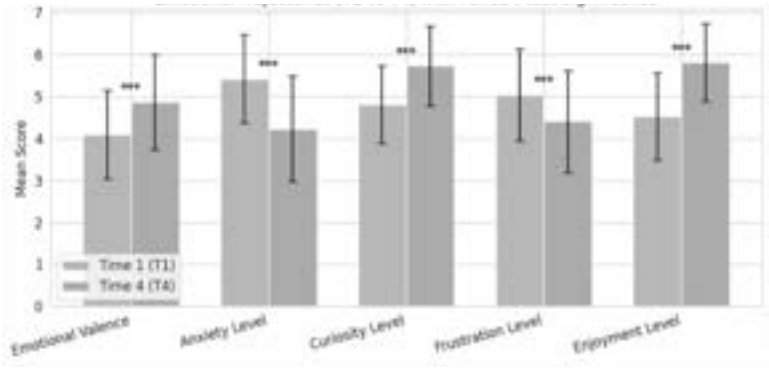


Figure 2. Emotional trajectories (T1 vs T4)with paired t-test significance

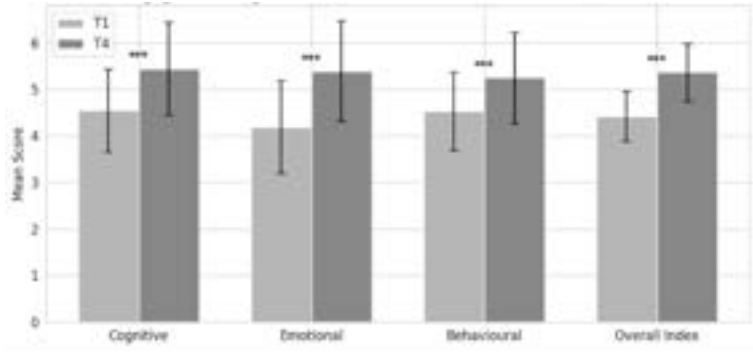


Figure 3. Engagement: cognitive, emotional, behavioural, and overall (T1 vs T4)

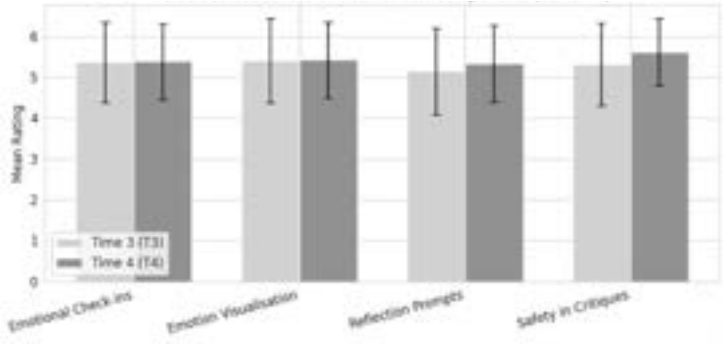


Figure 4. Perceived usefulness of the intervention components (T3 vs T4)

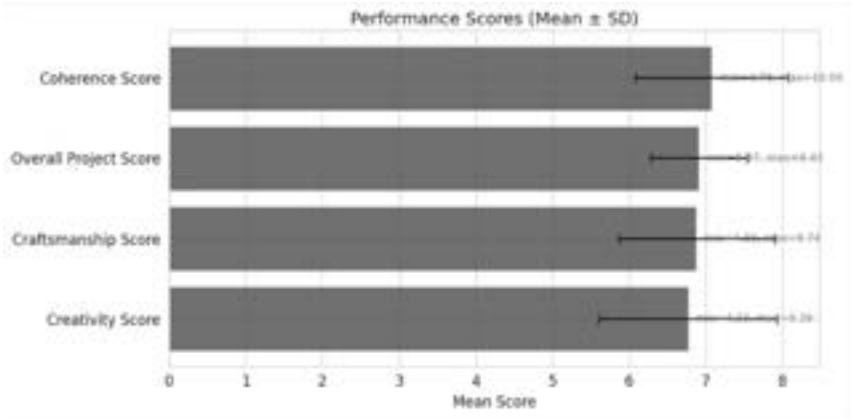


Figure 5. Performance scores

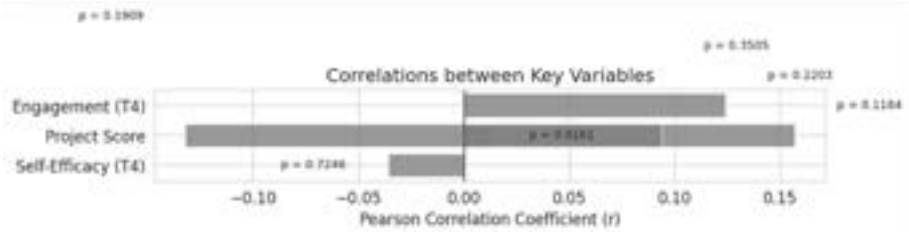


Figure 6. Correlations between key variables

5 DISCUSSION

5.1 Emotion-Scaffolding Pedagogy's Effectiveness on Student's Emotional State

Emotion-scaffolding pedagogy's effectiveness at countering the typical trend toward increasing levels of student distress as the deadline approaches is evidenced by the significant changes in students' emotional states from baseline to project completion. The results of this study suggest that Kansei-scaffolding practices could be contributing to an acceptance of emotional disclosure among students and providing them with the cognitive tools necessary to view discomfort as a normal and manageable part of creating something new instead of as an indicator of personal failure.

Of particular significance is the large increase in emotional granularity. A high level of emotional granularity has been linked to using more adaptive emotion regulation strategies, lessened amygdala activation, increased activity in the pre-frontal cortex, and greater overall cognitive flexibility (a key component of creative problem solving). This means that students will have a better understanding of when they are feeling overwhelmed, frustrated, or anxious and will be able to choose an appropriate way to deal with each state. By making explicit emotional literacy an equal partner in both aesthetics and technical competence when evaluating critiques, the author demonstrated that emotion can become more than a private, incidental element of learning and can be used as a valid criterion for making judgments regarding creativity. The increase in design self-efficacy is also relevant since creative self-efficacy has been shown to positively predict one’s ability to creatively perform, persist in working creatively, and desire to use divergent thinking. Given that there were separate improvements in design self-efficacy and ratings based upon the rubrics, there is clearly additional research that should be conducted to

determine if creative self-efficacy plays a role in process aspects of creative work (e.g., willingness to continue working on a difficult task, willingness to revise/iterate, and receptivity to receiving feedback) that may not be well-quantified by rubric-based evaluations of completed works.

Given prior studies showing that affective pedagogies improve emotional engagement, thereby sustaining cognitive involvement and behavioral commitment to the learning process, the sizable increases across all three aspects of studio engagement were consistent with these findings. Emotional Engagement experienced the greatest total change, suggesting that Kansei-scaffolding practices may have most effectively addressed a frequently overlooked area in traditional studio pedagogy: students' sense of connection to their own work, their peers, and the learning experience itself. Although moderate positive correlations existed between affective measures and creative outcomes, they highlight the need for a more comprehensive theoretical framework describing the relationship between emotion and design learning.

6 CONCLUSIONS

This research investigates the impact of integrating emotional intelligence training into design studio processes on students' emotional experiences and confidence during creative tasks. Results indicate significant improvements in positive emotions and decreases in negative feelings, suggesting that emotion-based teaching approaches may enhance well-being in design education. Although limitations exist, such as the lack of a comparison group and reliance on self-report data, findings imply that emotional scaffolding could positively influence the creative process. Future studies are proposed to explore the cumulative effects of emotion-scaffolding over time, and qualitative insights could clarify students' emotional understanding. The findings offer practical guidelines for design educators to incorporate emotional awareness into existing curricula, fostering a culture where discussing emotions is normalized in creative settings.

ACKNOWLEDGMENTS

The author thanks the undergraduate design students who participated in this study and the studio instructors whose collaboration made this emotion-scaffolded intervention possible

REFERENCES

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Boling, E., Gray, C., & Smith, K. (2020). Educating for design character in higher education: Challenges in studio pedagogy. In S. Boess, M. Cheung, & R. Cain (Eds.), *Synergy: DRS International Conference 2020, 11–14 August. Design Research Society.* https://doi.org/10.21606/drs.2020.120

Chuah, K. M., Chen, C. J., & Teh, C. S. (2008). Incorporating Kansei engineering in instructional design: Designing virtual reality based learning environments from a novel perspective. *Themes in Science and Technology Education*, 1(1), 37–48.

Choi, J. N. (2004). Individual and contextual predictors of creative performance: The mediating role of psychological processes. *Creativity Research Journal*, 16(2–3), 187–199.

Gray, C. M. (2013). Informal peer critique and the negotiation of habitus in a design studio*. Art, Design & Communication in Higher Education,* 12(2), 195–209.

Hammond, L., & Moore, W. M. (2018). Teachers taking up explicit instruction: The impact of a professional development and directive instructional coaching model. *Australian Journal of Teacher Education (Online)*, *43*(7), 110-133.

Ho, A. G. (2014). The new relationship between emotion and the design process for designers. *Archives of Design Research,* 27(1), 43–58. https://doi.org/10.15187/adr.2014.02.111.1.43

Ho, A. G., & Siu, K. W. M. (2012). Emotion design, emotional design, emotionalize design: A review on their relationships from a new perspective. *The Design Journal*, 15(1), 9–32. https://doi.org/10.2752/175630612X13192035508462

Jobst, B., & Meinel, C. (2012). How can creative self-efficacy be fostered in design education? In *Proceedings of the 14th International Conference on Engineering and Product Design Education* (pp. 1–6). Artesis University College.

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2018). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, 24(1), 10–16.

Mathisen, G. E., & Brønnick, K. S. (2009). Creative self-efficacy: An intervention study. *International Journal of Educational Research*, 48(1), 21–29. https://doi.org/10.1016/j.ijer.2009.02.009

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things.* Basic Books.

Park, M. H., Tiwari, A., & Neumann, J. W. (2019). Emotional scaffolding in early childhood education. *Educational Studies,* 46(5), 570–589. https://doi.org/10.1080/03055698.2019.1620692

Saarni, C. (2017). *The development of emotional competence*. Guilford Press.

Singaravel S, Suykens J, Janssen H, Geyer P. Explainable deep convolutional learning for intuitive model development by non–machine learning domain experts. *Design Science*. 2020;6:e23. https://doi.org/10.1017/dsj.2020.22.

ENHANCING HUMAN-PRODUCT INTERACTION PERFORMANCE: A KANSEI ENGINEERING APPROACH TO REDESIGNING THE TRADITIONAL TURKISH COFFEE POT

Ahmet Berat Akyol$*$ ^{a}

^{a} *Marmara University, Türkiye, akyolahmetberat@gmail.com*

ABSTRACT

Traditional craft products possess a strong cultural heritage, yet they often fall short of modern ergonomics and Human-Product Interaction (HPI) standards. This study presents an "Archetypal Optimization" model aimed at eliminating the negative emotional perceptions (Kansei) caused by the physical constraints of the traditional Turkish coffee pot (*cezve*). Within the scope of the research, 80 critical user feedback points focusing on design constraints across e-commerce platforms were analyzed using Thematic Analysis, and form-related complaints were mapped to negative Kansei adjectives (Tiring, Unsafe, Incompetent). To resolve these issues, parametric simulations were performed using Digital Human Modeling (DHM). The traditional 120° handle angle was found to force the wrist into ulnar deviation; by optimizing this axis to 145°, the RULA (Rapid Upper Limb Assessment) score was reduced from 6 to 2 (safe zone). Furthermore, a "U" shaped pouring spout design that resolves hydrodynamic issues was introduced, and haptic integrity was achieved through the material contrast between polished copper and solid wood. This study provides measurable evidence demonstrating how physical performance enhancements translate immediately into positive Kansei perceptions (Professional, Controlled, Safe) within the user's psychology.

Keywords: *Kansei Engineering, Archetypal Optimization, RULA, Human-Product Interaction, Turkish Coffee Pot.*

$*$ Ahmet Berat Akyol. Postal address: Marmara University, Department of Industrial Design, Acıbadem Mahallesi, Acıbadem Caddesi, No: 148/1, Uskudar, Istanbul, Türkiye. E-mail: akyolahmetberat@gmail.com. Phone: +90 545 382 9182.

1. INTRODUCTION AND PROBLEM DEFINITION

1.1. Traditional Context, Thermodynamics, and Invisible Labor

The traditional *cezve*, at the center of the Turkish coffee ritual inscribed on the UNESCO Representative List of the Intangible Cultural Heritage of Humanity, is a cultural archetype reflecting centuries of craft heritage and thermodynamic necessities with its wide base and narrow neck. Although electric models have automated the brewing process in recent years, the handle ergonomics of the product, the pouring action, and the emotional perception this classical form evokes in the user have been largely ignored. In particular, traditional stovetop use awaits an analytical optimization in terms of modern Human-Product Interaction (HPI) standards. However, in today's design approach, functional performance is not the only expectation from such objects carrying the "invisible labor" of artisans; users now seek haptic satisfaction, physical safety, and cognitive comfort within this experience.

1.2. Physical and Biomechanical Problem Definition

An examination of existing *cezve* typologies reveals that product geometry creates two fundamental engineering and ergonomic problems that directly undermine the user experience:

Spout and Flow Instability: Sharp "V"-shaped pouring spouts, a constraint of traditional manufacturing, disrupt the fluid dynamics during the transfer of coffee into the cup. The liquid sticking to the metal body and trickling down due to surface tension (Coanda effect) and the turbulence created by the narrow cross-section lead to both splashing and the loss of *crema*, the most valuable part of the ritual (Keller, 1957).

Postural Fatigue and Asymmetry: In standard designs, attaching the handle to the body at an angle close to 120°, combined with a shifting center of mass during pouring, forces the user's wrist into severe pronation and ulnar deviation. This postural deviation, corresponding to high risk scores in RULA analyses, causes acute fatigue (McAtamney & Corlett, 1993), while the fact that designs generally only allow right-handed use excludes universal design principles.

1.3. Research Objective and Scope

Physical and biomechanical deficiencies directly damage the emotional bond between the user and the product. Instead of radically altering the traditional *cezve* typology, this study aims to optimize the existing archetype through Kansei Engineering principles. Eliminating negative Kansei perceptions such as "Tiring", "Unsafe", and "Incompetent" distilled from consumer complaints is the primary objective of this research. Accordingly, a novel form synthesis (Nagamachi, 1995) has been developed to evoke feelings of "Safe", "Fluent", and "Professional" in the user by utilizing a wide "U" pouring spout that stabilizes fluid dynamics, a 145° biomechanical handle angle, and a copper-wood material contrast.

2. LITERATURE REVIEW

2.1. Kansei Engineering and Human-Product Interaction (HPI)

Kansei Engineering (KE) is a proactive methodology that transforms a user's abstract psychological responses to a product into concrete design parameters (Nagamachi, 1995). Schütte (2005) defines KE as a bridge linking abstract emotional states like "safe" or "tiring" to mathematical forms, geometric curves, and material decisions. Optimizing grip comfort and reducing the cognitive strain caused by the pouring action in cultural objects directly manipulated by hand, such as the *cezve*, directly enhances the emotional quality of Human-Product Interaction (HPI).

2.2. Modernization of Traditional Craft and Archetypal Optimization

In designs carrying cultural heritage, preserving product DNA (the archetype) is critical for maintaining the user's sense of belonging (Moalosi et al., 2010). With its centuries-old history, the *cezve* is a successful archetype that has already resolved functional dynamics such as heat transfer and foam preservation. Therefore, rather than alienating these deep-rooted forms through radical innovation, it is necessary to improve their ergonomic and sensory weaknesses using the "Archetypal Optimization" approach. When supported by modern hydrodynamic and biomechanical validations, this heritage typology carrying the invisible labor of artisans transforms into a high value-added contemporary design object.

2.3. Biomechanical Performance and Haptic Perception

In manipulative tools, the physical form determines the user's biomechanical health, while material selection dictates the tactile (haptic) experience. Incorrect angles in handled utensils deflect the wrist from its neutral axis (*ulnar/radial deviation*), leading to acute musculoskeletal fatigue (McAtamney & Corlett, 1993). Beyond physical ergonomics, material contrasts also influence Kansei perception. Experiential material combinations, such as smooth-cold metal and textured-warm wood, create feelings of trust, craftsmanship, and quality in the user (Karana et al., 2014). In this context, it is evident that the uniform metal texture and the body-straining handle angle of the traditional *cezve* must undergo a biomechanical and haptic optimization.

3. METHODOLOGY AND DATA ANALYSIS

3.1. Data Universe and Sample Selection

This research aims to map the relationship between the physical form of the traditional *cezve* and the user's sensory perceptions (Kansei). The data universe of the study consists of customer experiences compiled from the three most popular e-commerce platforms in Türkiye. To detect existing design flaws in the market, feedback from the last 24 months was scanned, and 80 critical data points were filtered. Since high-rated reviews in the Kansei Engineering perspective often contain expressions too abstract to be translated into design, only 1-to-3-star feedback focusing on moments of *functional frustration* experienced directly with the product geometry was included in the analysis.

3.2. Thematic Coding and User Voice

The collected qualitative data were coded using the Thematic Analysis method to provide a systematic structure to complex everyday language patterns (Braun & Clarke, 2006). The form architecture of the *cezve* was divided into five main themes: handle, neck, spout, body, and material. Users' daily complaints were directly translated into industrial design and engineering

terminology (*Coanda effect, ulnar deviation, thermal conductivity*). Some critical user experiences isolated from the 80-data pool are as follows:

Handle Ergonomics (Ulnar Deviation): "The handle angle is so steep... I have to bend my wrist excessively just to pour the coffee. It feels both unstable and tires my wrist a lot." (Data #05)

Spout Hydrodynamics (Coanda Effect): "The pouring spout is too narrow... The liquid constantly trickles down the body, and the foam is wasted before reaching the cup." (Data #36)

Body Form (Geometric Constraint): "The base is very wide but the neck is nonsensically narrow. It is impossible to clean the inside with a sponge; coffee grounds always remain in the corners." (Data #22)

Material and Haptic Perception (Kansei): "Its metal is thin as paper and the grip handle is slippery. It gives no sense of security when I hold it; it feels cheap and dangerous." (Data #11)

3.3. Kansei Inference and Semantic Mapping

Following the thematic analysis, the negative emotions evoked in the user's psychology by the identified technical flaws were mapped with reference to the Semantic Differential (SD) scale, which forms the basis of Kansei measurement (Osgood et al., 1957). The negative Kansei adjectives triggered by these technical design errors extracted from user experiences were classified according to their frequencies. This strategic mapping process revealed the foundational design guideline matrix (**Table 1**), which clarifies the positive emotions ("Safe", "Fluent", "Professional") that the newly designed product form should target during the Archetypal Optimization phase.

Table 1. Thematic analysis and Kansei Engineering mapping of traditional *cezve* user complaints (N=80).

Form Component	Frequency	Dominant Technical/Ergonomic Flaws	Evoked Negative Kansei	Optimization Target (Positive Kansei)
Handle	41%	High Thermal Conductivity, Static Instability, Weak Mechanical Connection, Incorrect Axis Angle (Ulnar Deviation)	Anxious, Unsafe, Tiring, Exclusive	Safe, Ergonomic, Balanced, Inclusive
Spout	27%	Surface Tension Flaws (Coanda Effect), Disrupted Laminar Flow, Foam Distribution Flaw	Uncontrolled, Incompetent, Messy	Fluent, Precise, Professional, Clean
Body	14%	Insufficient Base Diameter, Lack of Radius in Internal Geometry (Grounds Accumulation), Mass Distribution Flaw	Bulky, Laborious, Unstable	Practical, Balanced, Fitted, Dynamic

Material	12%	Chemical Leaching (Metallic Taste/Odor), Coating Instability, Low Thermal Shock Resistance	Toxic, Poor Quality, Flimsy	Healthy, Craftsmanship, Premium, Pure
Volume	6%	Capacity Illusion (Misleading Volume), Incorrect Anthropometric Form	Misleading, Dangerous	Transparent, Controlled, Honest

4. PROPOSED DESIGN SYNTHESIS

4.1. Parametric Form Exploration, Biomechanical Optimization, and RULA Analysis

Rather than creating a radical and culturally alien rupture in the traditional *cezve* form, an "Archetypal Optimization" process was conducted based on measurable engineering criteria. In line with Norman's (2004) behavioral design principles, the final geometry of the form was shaped through successive kinematic simulations. Handle Axis Iteration (120° to 145° Transition): To resolve the ulnar deviation problem caused by the steep handle angle of standard *cezves* on the user's wrist, the handle axis was analyzed parametrically. To eliminate human-induced error margins in physical testing, a Digital Human Modeling (DHM) approach was preferred. When the pouring action was simulated with the traditional 120° angle, shoulder flexion (*Upper Arm*) spiked into the 20°–45° range, triggering postural risk, while the wrist bending angle deviated by more than 15° from the neutral axis, causing severe ulnar deviation. When the handle axis reached 145°, shoulder flexion decreased to the 0°–20° range, which is optimal for human anatomy, wrist rotation was zeroed to a neutral position, and the center of mass achieved a perfect balance. The RULA sub-score distributions resulting from these parametric tests, presented in **Figure 1**, clearly demonstrate the direct impact of handle angle modification on body mechanics. By evolving the angle from 120° to 145°, joint loads were reduced to minimum levels, eliminating the feelings of "Tiring" and "Unsafe" to build a "Controlled" and "Professional" Kansei perception.

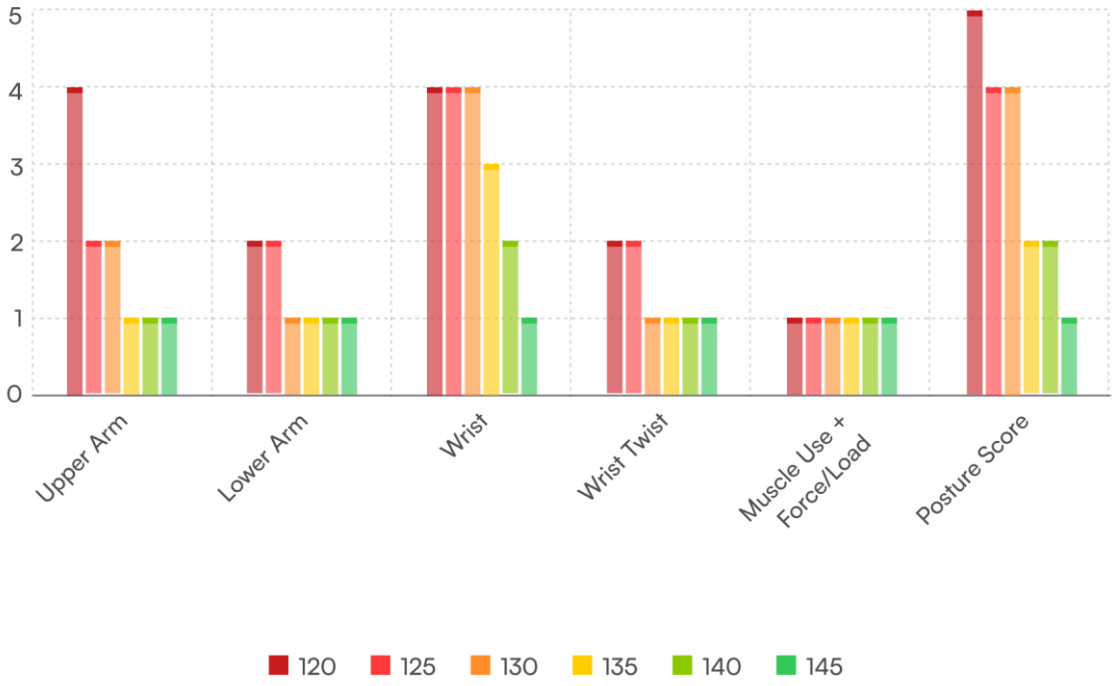


Figure 1. RULA sub-score distributions and postural risk analysis across different *cezve* handle angle iterations.

4.2. Spout Hydrodynamics and Flow Optimization

Transition from Sharp "V" Form to Teardrop "U" Form: The sharp "V"-shaped spout structure, a byproduct of traditional manufacturing methods, creates surface tension during pouring, causing the liquid to adhere to the body (known as the Coanda effect) (Keller, 1957). To solve this hydrodynamic problem, a wide and organic "U" (teardrop) spout geometry was designed to clarify the fluid's break-off point. This wide cross-section allows the coffee to glide into the cup via a laminar flow without entering turbulence, thereby preserving the traditional foam (*crema*) texture. Consequently, the negative Kansei perceptions of "Unsafe" and "Incompetent" triggered by uncontrolled pouring are transformed into the "Fluent" and "Precise" feelings shown in **Figure 4**, thanks to the hydrodynamic stability of the form.



Figure 2. Final design concept and optimized fluid dynamics of the proposed pouring spout.

4.3. Body Geometry

Sharp base-to-neck transitions in traditional *cezves* both disrupt heat circulation and cause coffee grounds to accumulate in the bottom corners, creating a functional barrier to cleaning. In the proposed design, these rigid transitions are resolved with a tectonic and organic curve featuring G2/G3 surface continuity (*curvature continuity*). This holistic flow within the form erases the "Bulky" and "Laborious" perceptions evoked by the traditional typology, giving the product a "Holistic" and "Ergonomic" Kansei value embodied in **Figure 3.**



Figure 3. Surface continuity (G2/G3) and holistic body geometry of the optimized design.

4.4. Material Selection and Haptic Kansei Perception

To preserve the cultural DNA of the product, the traditional copper and wood material palette was maintained, but optimizations were made to improve the tactile (haptic) and thermal user experience (Ashby & Johnson, 2013). Body (Polished Copper): Chosen to reflect cultural DNA, the copper body prevents temperature fluctuations with its high thermal conductivity, supporting the traditional slow brewing ritual. The high-gloss polished surface finish emphasizes the "Pure" character of the material, offering the user a refined craft experience. Handle (Walnut Wood): To counter the high thermal conductivity of copper, a brass ring (*heat break*) and solid walnut have been integrated into the body-handle interface. While the open-pore structure of the wood increases grip security (*affordance*), it generates the "Craftsmanship" and "Premium" Kansei perceptions targeted in **Table 1** through the material contrast showcased in **Figure 2.**



Figure 4. Material execution, haptic contrast, and the brass heat-break interface.

5. DISCUSSION

This study demonstrates that Kansei Engineering is not merely an aesthetic stylization tool; rather, it offers an analytical approach that guides user psychology in solving biomechanical and hydrodynamic problems. While Kansei applications in the literature are generally concentrated in modern sectors such as automotive or electronics (Nagamachi, 1995), this research contributes a unique context to the literature by focusing on the ergonomic modernization of a traditional cultural object (the *cezve*).

The research findings demonstrate the direct impact of physical performance on emotional perception. The perceptions of "Tiring" and "Unsafe" identified in **Table 1** are not related to the product's visual appearance alone; they are directly linked to the ulnar deviation caused by the traditional 120° handle. The radical drop in RULA scores achieved through the 145° axis optimization (**Figure 1**) validates how removing physical constraints immediately triggers the

feelings of trust and control that Norman (2004) conceptualizes at the *behavioral level*. At the exact point where physical fatigue is reduced, a "Professional" and "Controlled" Kansei perception begins.

Furthermore, arguments asserting that cultural DNA must be preserved (Moalosi et al., 2010) are put into practice through the "Archetypal Optimization" applied in this study. Instead of forcing the product into an alien or overly futuristic form, the familiar archetype was maintained using seamless G2/G3 surface continuity and a "U"-shaped pouring spout. Material perception, as emphasized by Ashby and Johnson (2013), was constructed through the haptic contrast between the thermal quality of polished copper and the open-pore texture of solid wood. This synthesis combines the user's expectation for a "Craftsmanship" and "Premium" product with the flow stability brought by modern engineering.

5.1. Limitations of the Study and Future Research Directions

The primary limitation of this study is that ulnar deviation and biomechanical tests were conducted using Digital Human Modeling (DHM) simulations rather than physical prototypes. Although simulations provide highly reliable anthropometric data, physical interaction is essential to measure actual thermal perception and haptic feedback during use. In future studies, it is planned to produce functional prototypes of the proposed design to conduct muscle activation tests on users via Electromyography (EMG) and to validate Kansei outcomes in a laboratory environment through Semantic Differential questionnaires.

6. CONCLUSION

This research has mapped the negative Kansei reflections in user psychology caused by structural and hydrodynamic constraints in the morphology of the traditional Turkish coffee *cezve*, and has optimized these limitations through measurable design parameters. In a process that began with the thematic analysis of 80 users, the postural risk score (RULA) was reduced from 6 to 2 by evolving the traditional 120° handle angle to 145°. Pouring control was enhanced through a laminar-flow-oriented "U" (teardrop) form replacing the sharp "V" spout structure, while the targeted perceptions of "Craftsmanship" and "Premium" were materialized through seamless surface continuity in the body and haptic contrast in the material palette.

Consequently, this study offers a methodological model demonstrating how archetypal objects identified with cultural heritage can be transformed using modern biomechanical and sensory design (Kansei) principles without losing their authentic identity. The developed "Archetypal Optimization" approach is envisioned to guide the modernization processes of other traditional objects that establish physical contact with humans in the future.

REFERENCES

Ashby, M. F., & Johnson, K. (2013). *Materials and design: the art and science of material selection in product design* (3rd ed.). Butterworth-Heinemann.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Karana, E., Pedgley, O., & Rognoli, V. (Eds.). (2014). *Materials experience: fundamentals of materials and design*. Butterworth-Heinemann.

Keller, J. B. (1957). Teapot effect. *Journal of Applied Physics*, 28(8), 859-864. https://doi.org/10.1063/1.1722880

McAtamney, L., & Corlett, E. N. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2), 91-99. https://doi.org/10.1016/0003-6870(93)90080-S

Moalosi, R., Popovic, V., & Hickling-Hudson, A. (2010). Culture-orientated product design. *International Journal of Technology and Design Education*, 20(2), 175-190. https://doi.org/10.1007/s10798-008-9077-1

Nagamachi, M. (1995). Kansei engineering: a new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15(1), 3-11. https://doi.org/10.1016/0169-8141(94)00052-5

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic Books.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

Schütte, S. (2005). *Engineering emotional values in product design: Kansei engineering in development* (Linköping Studies in Science and Technology. Dissertations, No. 951). Linköping University Electronic Press.

EXPLORING THE INTERACTION BETWEEN GENERATIVE AI AND KANSEI DESIGN PROCESSES

Feichen Wu[1a], Chun-Heng Ho[a]

[1a] National Cheng Kung University, Taiwan, P38133018@gs.ncku.edu.tw
[a] National Cheng Kung University, Taiwan, hoch@gs.ncku.edu.tw

ABSTRACT

Under the wave of the modern "experience economy," the focus of industrial design has shifted from functional satisfaction to experience shaping. The success of a product highly depends on its ability to precisely induce positive emotions in users, making "Kansei design" the core of contemporary design. However, with the widespread adoption of generative artificial intelligence (GenAI), its inherent lack of authentic emotion and empathy has raised academic concerns regarding potential negative interference when applied to such tasks that heavily rely on emotional connections. This study aims to investigate whether the intervention of generative AI leads to a deterioration in the quality of final Kansei design outcomes. Adopting a primarily qualitative case-controlled design, this study arranged for two professional designers with highly similar backgrounds to execute a Kansei design task for a "companion robot for children with autism" under "AI-assisted" and "non-AI-assisted" conditions. Cognitive dynamics were captured via the think-aloud method, and a retrospective analysis was conducted by cross-validating the think-aloud data with double-blind expert evaluation results. The results indicate that the design outcomes from the AI-assisted group performed better in emotional connection and Kansei quality. By providing domain knowledge, AI successfully served as an "empathy navigator," helping designers transcend cognitive blind spots, while the critical form-giving decisions that determined product approachability remained dominated by human intuitive aesthetics.

Keywords: *Kansei design, Generative AI, Human-AI collaboration, Think-aloud*

1 INTRODUCTION

At the beginning of the 21st century, traditional markets driven by functional satisfaction became saturated, accompanied by severe technological homogenization. The concept of the "Experience Economy," proposed by Pine and Gilmore (1998), swept the globe, prompting industrial design to officially shift its focus from a "service economy" to "staging experiences."

[1] Corresponding author.

This transition fully highlights the importance of Kansei design. Today, the success of a product depends not only on the novelty of its physical attributes, such as function or form, but more importantly, on its ability to accurately evoke positive emotions in users (Nagamachi, 1999; Norman, 2005).As the wave of Generative Artificial Intelligence (GenAI) sweeps across the globe, its ability to generate contextually coherent content based on user inputs and support highly sustained interactions through real-time responses (Firat, 2023) has been proven to increase the quantity and novelty of ideas, as well as expand the solution space in the design field (Bouschery, Blazevic, & Piller, 2023; Chandrasekera, Hosseini, & Perera, 2025). Despite AI's immense potential in enhancing design novelty, its inherent lack of emotional understanding becomes an unavoidable challenge in current human-AI collaborative design, particularly when dealing with design cores that heavily rely on emotional connections.

From the perspective of computational linguistics, Bender, Gebru, McMillan-Major, and Shmitchell (2021) explicitly point out that large language models are inherently products of probabilistic prediction based on massive datasets. They act like "Stochastic Parrots"—while their outputs possess the "form" of language, they entirely lack genuine understanding, meaning-making, and communicative intent regarding the real world. This is because their text generation is not grounded in communicative intent or the psychological state of the user, nor does the machine itself possess the ability to share thoughts with a listener.From a psychological perspective, Turkle (2011) also notes in her book Alone Together: Why We Expect More from Technology and Less from Each Other that the care displayed by robots or AI is essentially a superficial performance by algorithms: "Simulated thinking might be thinking, but simulated feeling is not feeling; simulated love is never love." Therefore, AI inherently lacks empathy and certainly cannot genuinely empathize with human emotions.

Based on the aforementioned limitation of AI's lack of empathy, it is easy to intuitively assume that introducing AI into "Kansei design" tasks—which require a high degree of resonance—will inevitably interfere with designers, steering them in a negative direction that fails to adequately consider the users' authentic emotions.

However, in actual design practice, AI does not produce outcomes independently but serves as a collaborative tool for human designers. To explore its impact on design outcomes, it is necessary to further clarify the timing of its intervention. If AI only intervenes in the later stages of design (e.g., rendering techniques, technical execution, or visualization), the emotional core of the design has already been established by the human designer, resulting in relatively limited interference with Kansei design. Nevertheless, research indicates that the current intervention of generative AI primarily occurs in the concept generation phase (Chen et al., 2025). At the dawn of conceptualization, when AI—which fundamentally lacks real emotion—is deeply involved in design ideation, will its highly rational output lead to a design outcome lacking Kansei qualities? This remains unclear in current human-AI collaboration research.Therefore, the research question of this study is: When AI intervenes in designers' Kansei design tasks, does it truly exert a negative impact on the final design outcome?

2 METHOD

This study aims to investigate whether the intervention of generative AI brings negative interference to the final outcomes of designers executing Kansei design tasks. Using a single Kansei design topic as the assignment, this study compares the differences in the final design outcomes of designers under two conditions: "AI-assisted" and "non-AI-assisted." To deeply dissect the cognitive black box of collaboration between human designers and AI, this study adopts a case-controlled approach primarily based on qualitative analysis, utilizing the think-aloud method to retrace and examine the process of design generation.

2.1 Participant Criteria

Participant Criteria To minimize interference in the design process caused by variations in individual professional backgrounds, design experience, or cognitive styles, this study recruited two designers with highly similar backgrounds to participate in the experiment. The demographic profiles of the two participants and their respective experimental condition assignments are shown in Table 1.

Table 1. Participant Background and Experimental Task Conditions

	Participant Background	Experimental Task Conditions
AI Group	Graduated with a Master's degree from the Department of Industrial Design, National Cheng Kung University, with an undergraduate major in the same discipline (Industrial Design), and possesses six months of relevant professional experience in industrial design.	GPT assistance was permitted across all stages of the design process, with the sole exception of the "sketching phase," during which the use of GPT for image generation was prohibited.
Non-AI Group	Graduated with a Master's degree from the Department of Industrial Design, National Cheng Kung University, with an undergraduate major in the same discipline (Industrial Design), and possesses six months of relevant professional experience in industrial design.	No generative AI tools were permitted at any point throughout the entire design process.

2.2 Experimental Procedure

Both participants were required to independently complete a Kansei design task under identical constraints and environmental conditions: "designing a companion robot for children with autism." Within a 60-minute timeframe, the participants navigated the stages of problem understanding, concept development, and solution ideation, ultimately producing a conceptual sketch. The data collection and experimental procedure of this study were systematically divided into the following four core stages:

Task Description and Think-Aloud Practice: Prior to the experiment, the researcher provided a detailed briefing on the design topic and deliverable specifications. A think-aloud practice session was also conducted, during which participants were required to continuously verbalize their real-time thought processes throughout the design task.

Design Execution and Think-Aloud: Participants commenced the design task according to their assigned condition (AI group or non-AI group). The entire design process was documented via full audio and screen recording to capture real-time cognitive dynamics. Upon completion of the task, the audio recordings were fully transcribed into "think-aloud transcripts," serving as the raw textual data for subsequently tracing the designers' cognitive threads.

Sketch Style Standardization: To prevent external visual styles—such as individual sketching techniques or aesthetic line-work qualities—from interfering with the subsequent expert evaluation, the researcher utilized AI post-experiment to standardize the rendering style of both sketches.

Expert Evaluation and Interview: The two standardized conceptual sketches were presented to an expert—a PhD candidate in industrial design with professional teaching experience—for comparative evaluation to determine which concept demonstrated superior design quality. Following the evaluation, the researcher conducted a semi-structured interview with the expert (e.g., "Which concept performed better overall? What were the reasons? Why was the other concept less ideal?"). This aimed to deeply excavate the critical design features that influenced the expert's decision. The interview was fully audio-recorded and transcribed into "expert interview transcripts," establishing the data foundation for subsequent causal analysis.

2.3 Data Analysis

The data analysis in this study aims to elucidate the causal relationship between the "quality of the final Kansei design" and "AI intervention." The study adopts an outcome-oriented, retrospective analysis that cross-validates the expert evaluation data against the think-aloud process data. The specific analysis workflow is illustrated in Figure 1: Data Analysis Process Diagram.

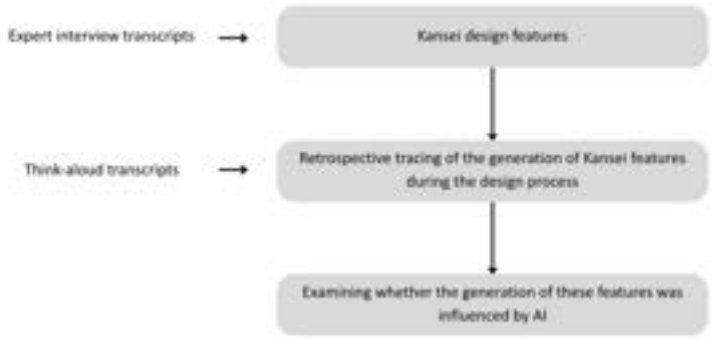


Figure 1. Data Analysis Process Diagram

First, the "expert interview transcripts" are analyzed. From the expert's evaluations, specific design features and conceptual details that led her to perceive a concept as superior are extracted. Next, using the design features identified by the expert as analytical anchors, the analysis traces back to the corresponding participant's (either from the AI or non-AI group) "think-aloud transcripts" to locate the critical cognitive segments where that specific feature was generated and developed during the design process. Finally, an analysis of the cognitive context

within these critical segments is performed to examine the source of inspiration for the design feature. This allows the researcher to retroactively deduce the underlying cause of the feature: whether it stemmed from the designer's pure projection of personal experience, or was actively stimulated by the information provided by the AI. Conversely, if the expert notes that a certain design failed to adequately address the user's emotions, the analysis similarly traces back to examine whether this deficiency resulted from the negative interference of the AI's formulaic output.

3 RESULTS

During the quality evaluation phase of conceptual sketches, under group-blind scoring conditions, the expert determined that the design solution from the "AI group" performed better in terms of overall design quality. According to the expert interview transcripts, the expert explicitly stated that three design features in the AI group's design (Figure 2) were the primary factors driving the overall design quality evaluation, namely: the "healing plush material," the "endearing 'ugly-cute' appearance," and the "core device compatible with various plush toys."

Notably, in the evaluation of manufacturing and mass-production feasibility, the expert considered both solutions to possess a high degree of maturity. The expert emphasized in the interview: "In fact, the feasibility of implementing this [Figure 3, non-AI group] is extremely high. Its molding... whether for the facial expressions or the camera component, is actually very easy to achieve. It can be put this way—both of them are definitely feasible." Having ruled out the technical barriers of physical manufacturing, the core reason the expert awarded a higher evaluation to the AI group's solution primarily rests on the fact that the aforementioned three design features could more precisely align with the specific behavioral habits and emotional needs of children with autism.



Figure 2. AI Group Design



Figure 3. Non-AI Group Design

3.1 Design Feature 1: Healing Plush Material

In the expert interview, the expert repeatedly noted that the "plush" material in the AI group's solution served as a significant basis for their preference. The expert explicitly stated: *"Usually,*

children prefer this kind of plush toy or material because it feels a bit softer." The expert further explained that the presentation of the plush material could convey warmth and enhance the sense of companionship, thereby making children with autism more willing to approach and accept it. By contrast, the companion robot presented in the sketches of the non-AI group resembled a generic companion robot, which felt relatively colder and was comparatively deficient in establishing an emotional connection.

Retrospective analysis of the think-aloud process of the participant in the AI group reveals that the generation of this feature is directly correlated with the contextual information provided by the AI. Transcript analysis indicates that during the early stage of concept development, the participant received background descriptions from the AI regarding *"multi-sensory experiences"* for children with autism, which specifically highlighted the soothing role of touch in companionship contexts. Stimulated by this information, the participant stated: *"Basically, I'm thinking, what I want to make right now is something like a plush toy, because I think it mentioned touch just now. This robot itself will definitely have vision and hearing—it will definitely have sound—and then adding touch will make it feel like a healing toy."*

3.2 Feature 2: Core Device Compatible with Various Plush Toys

The expert also particularly emphasized that the concept of *"separating the core device from the outer shell"* in the AI group's design possessed a high degree of creativity. The expert pointed out that this design approach allows the core functions to remain consistent while the external appearance can be replaced according to needs; this type of customization-like design strategy is uncommon among similar companion products, yet it perfectly aligns with the specific context of autistic children who often exhibit *"fixed preferences and a strong sense of routine."*

Comparing this with the participant's think-aloud transcripts, this design feature also originated from the guidance of AI information. In the conversational context, the AI mentioned that *"repetition and regularity equal a sense of security"* and that *"children with autism may have their own specific obsessions."* Upon receiving these keywords, the participant verbally deduced: *"Because, just going back to what was said earlier, children with autism might have more of their own principles, and the things they like are very fixed and they're a bit obsessed with them... If it's a plush toy or some doll they liked before, I feel like if you suddenly introduce a companion robot to them today, they might not necessarily be able to accept it—that's not a certainty. So the fastest way to get them to touch and use this product, I think, is to directly attach it to something they already really like, and then they might feel, oh, this thing has come alive.*"

3.3 Design Feature 3: Endearing 'Ugly-Cute' Appearance

The third design feature pointed out by the expert in the AI group's design was "asymmetry" (specifically the design with one large ear and one small ear). The expert believed that: *"It's this image... because it's originally facing children with autism. So children with autism should relatively, you know, need something with more warmth... Although this thing looks a bit ugly-cute, it can actually convey that kind of emotional, you know, feeling—that kind of feeling of emotional exchange."* This asymmetrical, somewhat *"ugly-cute,"* and less conventional presentation broke the icy feeling of standardized products; conversely, the design of the non-AI

group was overly symmetrical and regular, making it easily appear rigid and creating a distance from children.

However, the emergence of this feature was likely not influenced by AI. When this study used this feature as an anchor to trace back through the think-aloud transcripts of the AI group participant, no record was found regarding the deductive process of external asymmetry or any influence from AI prompt information. The transcripts show that when drawing the ears, the participant only briefly stated verbally: *"Okay, then first, let's draw a rabbit...",* and immediately proceeded to draw, without consulting the AI or parroting any AI suggestions during the process. This result indicates that the design feature of the *"ugly-cute appearance"* was not an effect of AI intervention, but rather stemmed from the participant's individual aesthetic intuition and design style.

4 DISCUSSION

This study aims to investigate whether the intervention of generative AI in Kansei design tasks, which heavily rely on emotional connections, exerts negative interference on the final design outcomes. Through a comparative analysis utilizing the think-aloud method and single-blind expert evaluation, the results of this case study indicate that the intervention of generative AI did not lead to a deterioration in the quality of Kansei design as expected. On the contrary, within specific contexts, AI assistance facilitates designers in deepening their understanding of the behavioral characteristics of unfamiliar user groups.

The research findings show that the expert evaluation deemed the design solution from the AI-assisted group to possess superior Kansei design quality. Based on this result, the following section will delve into the role and actual impact of AI intervention during the Kansei design process.

4.1 The Correlation Between Domain Knowledge Provided by Generative AI and the Translation of Kansei Design Features

Although previous research, such as Bender et al. (2021), has pointed out that large language models are essentially "stochastic parrots" lacking real-world understanding, and Turkle (2011) questioned whether simulated emotions by machines can truly establish genuine connections with humans, the results of this study overturn the intuitive assumption that "the intervention of emotionless AI will inevitably lead to the deterioration of Kansei design outcomes."

Expert evaluation results indicate that the AI-assisted group outperformed the non-AI group in the quality of establishing emotional connections. Retrospective analysis of the think-aloud transcripts confirms that this advantage stems from a clear cognitive translation path: First, generative AI provides the designer with objective domain knowledge about the specific user group. Second, upon receiving such specialized knowledge, the designer interprets it through their own empathy and translates it into concrete design features. Finally, these features are accurately identified by the expert during the evaluation and deemed effective in establishing an emotional connection with autistic children. This specific process—from "AI providing

knowledge" to "the designer exercising empathy" and finally "translating into design features"—can be validated through the development of two features highlighted by the expert: the "healing plush material" and the "core device compatible with various plush toys."

First, regarding the generation of the "healing plush material," its emergence was not the result of the participant's random exploration. Rather, it originated from extracting descriptions about "multi-sensory experiences" and "the core role of touch in companionship contexts" from the background data provided by the AI. This objective knowledge successfully guided the designer to project empathy toward sensory pain points beyond the visual and auditory realms of autistic children, thereby transforming a conceptual technological device—originally leaning toward functional orientation—into a plush toy offering a warm sense of companionship.

Second, in the development of the "separation of the core device and outer shell," the AI provided information indicating that autistic children might have their own specific obsessions and a strong attachment to fixed routines or objects. This information further directed the designer's empathy toward the users' deeper psychological adaptation levels. Notably, although this critical knowledge was provided by the AI in the early ideation phase, the designer did not immediately generate a corresponding structural solution upon seeing the information. Instead, it was not until the mid-to-late stages of concept development, specifically during concrete form-giving decisions, that this pre-implanted knowledge was retrieved. During the think-aloud process, the designer noted that because autistic children are relatively stubborn about what they like, they might not accept a completely unfamiliar companion robot suddenly introduced to them.

Based on this understanding of user behavioral characteristics, and assisted by the background knowledge provided by the AI, the designer overcame their knowledge blind spots and translated them into a new countermeasure: extracting the core device housing the technical functions so it could be flexibly inserted into existing toys the child is already familiar with. This solution cleverly accommodates the obsessive traits of autistic children and mitigates the defensiveness and anxiety they might experience when encountering new products.

This study postulates that the role generative AI plays in Kansei design tasks is not as a "generator" of emotion, but as a "navigator" of empathy. While human designers possess empathy, they may struggle to focus it when facing unfamiliar user groups due to a lack of concrete understanding of user behavioral patterns. Although AI lacks emotion, the domain knowledge it provides serves as a directional guide for human empathy, pointing designers toward where their empathy can be most effectively applied.

Based on the postulated "navigation" value of AI in the aforementioned analysis, this study further infers that the intervention of generative AI offers a unique advantage in the early stages of Kansei design that traditional search engines cannot easily replace. As Radlinski and Craswell (2017) pointed out when exploring conversational search theory, traditional search engines are essentially one-way searches lacking systemic proactivity; the system does not actively offer guidance. Designers must iterate their searches, figuring out how to learn the correct vocabulary and keywords needed to accurately describe their information needs. In other words, if a designer cannot precisely find the right keywords to ask the question, the search engine cannot provide

the necessary information. However, in the early stages of Kansei design, when facing an unknown and specific user group, designers can easily fall into knowledge blind spots. For instance, if a designer is completely unaware of an autistic child's deep psychological need for "tactile soothing" or their "resistance to change" before conceptualization, a traditional search engine will not proactively compile this relevant information. This could lead to a critical issue: as long as the designers themselves do not anticipate these hidden emotional needs, they cannot acquire the relevant knowledge through traditional search, potentially resulting in missed opportunities to consider and respond to users' emotional needs throughout the design process. This may also account for why the participant in the non-AI group chose to style the product's appearance as a conventional, cute robot featuring a technological aesthetic.

4.2 The Impact of Generative AI Intervention on the Designer's Subjectivity

As discussed in the previous section, generative AI successfully guided the generation and development of design features such as the "healing plush material" and the "core device compatible with various plush toys" by providing domain knowledge in the Kansei design task. However, upon further examining another key feature raised in the expert evaluation—the "endearing 'ugly-cute' appearance"—this study observed a completely different cognitive decision-making path. Retrospective verification of the transcripts confirms that this unconventional form-giving feature was not influenced by any AI prompts during the concept development process; rather, it stemmed purely from the designer's intuitive aesthetics when considering autistic children. This decision-making phenomenon, independent of AI, brings up a frequently discussed question in the design field over the past two years: In a human-AI collaborative environment reliant on AI knowledge navigation, do human designers still possess irreplaceable value? That is, will AI eventually replace human designers?

This study argues that the feature of the "endearing 'ugly-cute' appearance," generated without AI influence, demonstrates that even in an AI-intervened design task environment, human designers retain their aesthetic judgment and decision-making power over form. In other words, while AI can provide knowledge and conceptual direction, the final determination of whether a product possesses approachability and positive emotional traits remains reliant on the designer's own aesthetic intuition and subjective judgment. This echoes the findings of Chen et al. (2025), which indicate that in collaboration with AI, AI primarily assists humans in the problem definition and idea generation phases, while the creative selection and evaluation phases are still predominantly human-led. Furthermore, Booth, Donohew, Wlezien, and Wu (2024) noted in a McKinsey report that although generative AI can significantly accelerate conceptual ideation, it is not an omnipotent magic wand; translating AI outputs into true design value still relies on the human designer's discretion. AI excels at rational tasks such as context analysis and professional knowledge organization; however, establishing an emotional connection between a product and a user through design features remains a task that current AI cannot independently accomplish.

Nevertheless, this study discovered in the verbal data of the AI-assisted group that, during the early stages of the design process, participants exhibited a passive tendency to unconditionally accept the massive volume of unfamiliar information output by the AI. In the verbal data, this manifested as blindly parroting the information provided by the AI without immediate feedback

or critical thinking. This phenomenon aligns with the "overreliance" observed by Buçinca, Malaya, and Gajos (2021), where users tend to directly accept AI-generated results, even when the information provided by the AI might be incorrect. To a certain extent, this phenomenon reflects a crisis of the "compromise of subjectivity" among humans during human-AI collaboration.

In conclusion, the intervention of generative AI does not diminish the importance of human designers; on the contrary, within the division of labor in human-AI collaboration, it precisely accentuates and underscores the core value of human intuitive aesthetics. AI functions more as a powerful knowledge repository, responsible for expanding and stabilizing the rational boundaries of the problem space, whereas human designers retain decision-making authority within the solution space, translating their own intuition, style, and empathy into design features with deeper emotional connections. In other words, the core value of generative AI in Kansei design does not lie in replacing the emotional judgment of human designers, but rather in assisting them to more effectively understand unfamiliar user groups; ultimately, it remains the human designers themselves who determine how emotional connections are concretely translated into product forms and design features.

5 CONCLUSION

This study aims to explore a core anxiety in the contemporary design field: when generative AI, which inherently lacks authentic emotion, intervenes in Kansei design tasks, does it exert negative interference on the final design outcomes? Through a comparative analysis utilizing the think-aloud method and double-blind expert evaluation, the results of this study indicate that the intervention of generative AI did not lead to a deterioration in Kansei design quality as expected; on the contrary, it enhanced the emotional connection of the final design by establishing a rational knowledge framework.

This study further reveals the unique collaborative mechanism played by generative AI in Kansei design. The findings confirm that the core value of AI lies in its role as an "empathy navigator." Specifically, when facing unfamiliar user groups, AI can proactively deliver precise medical and behavioral characteristics, helping designers transcend unknown cognitive blind spots and focus their empathy onto the correct pain points and needs. However, this does not imply a weakening of the human designer's subjectivity. This study also found that after clarifying specialized knowledge, the product appearance—which determines the final degree of approachability—is still dominated by the human designer. Nevertheless, the study also uncovered a potential crisis in the early stages of the AI-intervened design process: when confronted with a massive volume of unfamiliar information output by the AI, participants exhibited a tendency toward "overreliance."

Based on the aforementioned findings, this study proposes concrete recommendations for future Kansei design practice and education. Given that AI can effectively establish knowledge frameworks and thereby facilitate the generation of empathetic features, future design practice and education should focus on cultivating designers' ability to effectively translate AI-generated

information into design features. This will enable designers to more efficiently harness AI tools to execute precise and warm Kansei designs tailored to specific or unfamiliar user groups.

Finally, although this study deeply analyzes the cognitive processes of AI intervention through the think-aloud method, several research limitations remain. First, adopting a case-controlled design, this study primarily focuses on the deep exploration of qualitative processes; due to the small sample size, its generalizability for quantitative inference remains limited. Second, the design task in this study was restricted to a "companion robot for children with autism." This task itself possesses a distinct cross-disciplinary medical knowledge background, allowing the knowledge-provision function of AI to be fully leveraged. Future research is suggested to expand the sample size of participants or test other types of Kansei design tasks to investigate whether the impact of generative AI intervention on Kansei design performance remains consistent across different design contexts. Furthermore, whether the phenomenon of overreliance exhibited by users during the early stages of human-AI collaboration is exclusive to rapid design tasks remains a question that warrants further investigation through full-scale, comprehensive design projects. Based on the aforementioned limitations and a synthesis of the findings from this study, concrete recommendations can be proposed for future Kansei design practice and education. Given that AI can effectively establish a rational knowledge framework and thereby facilitate the generation of empathy-driven features, future design practice and education should emphasize cultivating designers' capability to effectively translate AI-generated information into design features. This will enable designers to harness AI tools more efficiently, allowing them to execute precise and warm Kansei design tailored to specific or unfamiliar user groups.

REFERENCES

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?ü¶ú.* Paper presented at the Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.

Booth, B., Donohew, J., Wlezien, C., & Wu, W. (2024). Generative AI fuels creative physical product design but is no magic wand. *McKinsey & Company website*.

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer‐based language models. *Journal of product innovation management, 40*(2), 139-153.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction, 5*(CSCW1), 1-21.

Chandrasekera, T., Hosseini, Z., & Perera, U. (2025). Can artificial intelligence support creativity in early design processes? *International Journal of Architectural Computing, 23*(1), 122-136. doi:10.1177/14780771241254637

Chen, L., Song, Y., Guo, J., Sun, L., Childs, P., & Yin, Y. (2025). How generative AI supports human in conceptual design. *Design Science, 11*, e9. doi:10.1017/dsj.2025.2

Firat, M. (2023). How chat GPT can transform autodidactic experiences and open education?

Nagamachi, M. (1999). Kansei engineering: A new consumer-oriented technology for product development. *The occupational ergonomics handbook, 33*, 1835-1848.

Norman, D. A. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Pine, B. J., & Gilmore, J. H. (1998). Welcome to the experience economy.

Radlinski, F., & Craswell, N. (2017). *A Theoretical Framework for Conversational Search*. Paper presented at the Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway. https://doi.org/10.1145/3020165.3020183

Turkle, S. (2011). *Alone Together: Why We Expect More From Technology and Less From Each Other*. New York: Basic Books.

FRACTAL HIERARCHY IN WILLIAM MORRIS'S DESIGNS: EXPLORING THE FACTORS BEHIND IMPROVED QUALITY OF LIFE FOR THE COMMON PEOPLE

Genta FURUSAWA,

Kyushu University Press, Japan, 21belmore@gmail.com

ABSTRACT

In this paper, building on the pioneering research on "fractal fluency" by Richard Taylor et al., the author examines the fractal properties of William Morris's designs, as well as the hierarchical structure and homogeneity across different levels. The methodology employed in this study involves image analysis using the box-counting method. The concept of "fractal" discussed in this paper refers to "statistical fractals" and is distinguished from "strict fractals" such as the "Mandelbrot set." Furthermore, drawing on Kaplan's "Attention Recovery Theory (ART)," the paper argues that Morris's fractal designs, which mimic natural growth processes, likely brought a new appreciation for nature to society and contributed to improving the quality of life for the common people of that era.

Keywords: fractal, box-counting method, Attention Recovery Theory

1 INTRODUCTION

This study aims to examine, through his own works, the impact that the textile designs created by William Morris had on improving the quality of life for people living in the late 19th century. Morris is often referred to as the "father of modern design" and was the figure who initiated the lineage of modern design—from the Arts and Crafts movement to the Bauhaus. The motivation for this study is to explore the source of the innovation he brought about. In this paper, we will examine the wallpaper he designed as one aspect of this inquiry.

Following the Industrial Revolution, wallpaper became mass-produced and widespread among the general public with the advent of the cylinder printing press (patent obtained by Potters & Ross in 1839). However, products of the time were often of poor quality, imitating realistic paintings, or featured simple geometric grid patterns tailored to the machinery; issues included visual fatigue caused by uneven composition and psychological oppression resulting from a "cold grid" Morris's work serves as an antithesis to these trends [1].

This study compares pre-Morris wallpaper with Morris's works using the concept of "fractal fluency" proposed by Taylor et al. (a perceptual characteristic in which a moderate fractal dimension, D = 1.3–1.5, reduces physiological stress in humans) [2].Note that this paper deals not with geometric fractals, but with random "statistical fractals" found in nature. The aim is to measure and verify the fractal dimensions of both and to clarify the underlying factors contributing to the comfort of Morris's designs in living environments.

2 METHODS AND SUBJECTS / THEORETICAL FRAMEWORK

The box-counting method was used to measure the fractal dimension (D-value) [3]. The work selected for analysis is Morris's "Willow Bough" (1884) (Fig. 2). The work chosen for comparison is a pre-Morris piece, "Floral Bouquet in Rococo Scroll Border" (circa 1850–1860) (Fig. 1).The latter is a work that depicts flowers such as roses and dahlias in a three-dimensional and realistic manner, as if in a painting, and frames them within a luxurious Rococo-style border. It is one of the classic examples that was harshly criticized at the time as a "design that forcibly places three-dimensional realistic paintings on a two-dimensional wall" and as "unnatural and in poor taste."



Figure 1. *Floral Bouquet in Rococo Scroll Border*, maker unknown, circa 1850–1860



Figure 2. *Willow Bough*, William Morris, 1887

We cropped each of these images into individual units intended for wallpaper patterns, and to ensure uniform conditions, we resized them into 1,024-pixel squares. We then performed multi-scale sampling by taking samples from seven locations on each image using squares with sides of 128px, 256px, and 512px, enabling us to measure objective changes resulting from variations in the cropped size.

Furthermore, to prevent errors caused by differences in line thickness, measurements were performed using automated skeletonization to standardize line thickness and automated image binarization to convert the image colors to black and white.

3 RESULTS AND ANALYSIS

The results of the image measurements based on the method described in the previous section are as follows.

Table 1. D-values for each scale and region in Figure. 1

Sampling frame	Loc 1	Loc 2	Loc 3	Loc 4	Loc 5	Loc 6	Loc 7
128 px	1.412	1.419	1.465	1.096	1.510	1.106	1.556
256 px	1.489	1.689	1.212	1.537	1.625	1.664	1.193
512 px	1.843	1.830	1.838	1.866	1.845	1.893	1.837

Table 2. D values for each scale/site in Figure. 2

Sampling frame	Loc 1	Loc 2	Loc 3	Loc 4	Loc 5	Loc 6	Loc 7
128 px	1.504	1.490	1.425	1.420	1.398	1.442	1.598
256 px	1.767	1.743	1.764	1.724	1.751	1.765	1.785
512 px	1.934	1.927	1.930	1.921	1.921	1.920	1.925

Table 3. Mean D values and standard deviations (σ) for each scale in the analysis

Sampling frame	Figure 1. Average D-value / (Standard Deviation)	Figure 2. Average D-value / (Standard Deviation)
128 px	1.3665 / (σ = 0.1740)	1.4681 / (σ = 0.0638)
256 px	1.4870 / (σ = 0.1909)	1.7569 / (σ = 0.0184)
512 px	1.8504 / (σ = 0.0204)	1.9255 / (σ = 0.0048)

We measured seven randomly selected 128-pixel sections from Figure 1. The results show that the lowest D-value was 1.096 and the highest was 1.556 (Table 1), with a standard deviation of 0.1740, indicating variability depending on the location of the section (Table 3). However, the average D-value was 1.3665, which falls within Taylor's range for fractal fluency. The standard deviation also showed significant variation for the 256px and 512px samples.

When we measured seven randomly cropped sections of Figure 2 at 128px, the lowest D-value was 1.398 and the highest was 1.598 (Table 2), but the standard deviation was 0.0638, indicating minimal variation (Table 3). Furthermore, the average D-value was 1.4681, which also falls within Taylor's range of fractal fluency. For 256px and 512px, it was found that as the image was enlarged, the D-value increased, approaching that of a two-dimensional plane, while the standard deviation remained stable with little variation.

4 DISCUSSION

The experimental results show that at the 128px scale (the distance at which a person approaches the wallpaper and recognizes the leaves), both cases met Taylor's comfort threshold (D = 1.3–1.5) in terms of their mean values. However, in Figure 1 (Flowers and Columns), the standard deviation (σ = 0.1740) is high, indicating significant variation across different locations. In contrast, Figure 2 (Morris) is stable across the entire area with σ = 0.0638, allowing observers to experience consistent comfort regardless of where they look. Furthermore, as the scale is enlarged (to 512px), Figure 2 converges to a two-dimensional plane with overwhelming uniformity, achieving σ = 0.0048.

This "multiscale duality" can be explained by environmental characteristics in Kaplan's Attention Recovery Theory (ART) [4][5].Whereas wallpaper of a previous era forced residents to expend active attention, Morris's "Willow" provides "fascination" that stimulates spontaneous human attention through local (128px) geometric fluctuations, while ensuring "extent"—harmony in the space—through the minimal standard deviation at the macro level (512px). In

other words, through his editorial design, Morris is believed to have created a “foundation for visual attention recovery” that alleviates the mental fatigue of urban dwellers, thereby contributing to the improvement of people’s quality of life.

5 CONCLUSION

It appears that fractal fluency is not achieved simply by having nature present. This is because Figure 1, used in the experiment, also depicted numerous flowers, which are natural elements. It becomes clear that what Morris did was, in essence, an act of curation and design: he first decomposed nature, identified its essence, and then restructured it to fit the context of the living environment.

This study did not directly conduct an affective response evaluation through subject experiments. However, it revealed that Morris’s products met the mathematical environmental benchmark of “fractal dimension 1.3–1.5”—a standard for which Taylor et al. have demonstrated physiological stress-reducing effects—across the entire design at a specific scale. This confirmed that, emerging from the poor visual environments of the Industrial Revolution, Morris’s designs gradually spread into the living spaces of the common people and were able to function as a foundation for daily life that brought about their physiological and mental harmony.

This paper is a comparative case study (preliminary examination) focusing on Morris’s “Willow” and a representative example of pictorial wallpaper patterns from an earlier era. The high standard deviation observed in the earlier sample mathematically suggests remnants of the pictorial composition of the 1840s.The control of multiscale structures demonstrated in this study is a manifestation of the curatorial process through which Morris reinterpreted the essence of nature for the living space; it can be said that this went beyond mere decoration to serve as a catalyst for bringing a new, harmonious understanding of society and nature to the common people of the time. A future task will be to further increase the number of samples analyzed and statistically verify the trends in average variance across different eras.

REFERENCES

[1] Morris, W. (1883). *Hopes and Fears for Art*. (3rd ed.). London, UK: Ellis & White.

[2] Taylor, R. P., & Spehar, B. (2016). Fractal Fluency: An Intimate Relationship Between the Brain and the Processing of Fractal Stimuli. In *The Fractal Geometry of the Brain* (pp. 485-496).

[3] Sarkar, N., & Chaudhuri, B.B. (1992). An efficient approach to estimate fractal dimension of textural images, *Pattern Recognition*, Volume 25, Issue 9 (pp. 1035-1041).

[4] Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective*. Cambridge, UK: Cambridge University Press.

[5] Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. In *Journal of Environmental Psychology*, 15(3) (pp. 169-182).

GENERATING TIRE TREAD DESIGNS USING A SENSORY EVALUATION REGRESSION MODEL OF GENERAL USERS AND TIRE MANUFACTURER EMPLOYEES AND A GENERATIVE MODEL

Ikki KODA[a], Ren HAYAKAWA[a] , Takeo KATO[1b]

[a] *Graduate School, Keio University, Yokohama, Japan*
[b] *Department of Mechanical Engineering, Keio University, Yokohama, Japan*

ABSTRACT

In this study, a CatBoost regression model was constructed to clarify the relationship between geometric features extracted from tire tread images and sensory evaluations of “quietness” and “grip performance” on DRY and WET surfaces. SHapley Additive exPlanations (SHAP) was then applied to visualize the contribution of each feature, enabling the identification of geometric features that are important for sensory evaluation. Furthermore, by comparing the sensory evaluations of general users and tire manufacturer employees, we identified the geometric features that caused differences in their evaluations. In addition, the evaluation values for CatBoost regression model were used to screen the training data input to StyleGAN, generating tire tread images likely to obtain high sensory evaluations of “quietness” and “grip performance” on DRY and WET surfaces from both general users and tire manufacturer employees.

Keywords: *Generative design, Tire tread, CatBoost, SHAP, StyleGAN*

1 INTRODUCTION

Tire treads affect major tire performances such as “quietness” and “grip performance” (Lee et al., 2022). Their appearance also shapes impressions of performance and influences consumers’ purchase intention (Kii et al., 2023; Creusen & Schoormans, 2005). Although many studies have examined the relationships between “quietness” and “grip performance” and geometric features of tire treads (Bridgestone, 2008; Kim & Jeong, 2010), few have investigated the relationship between tire tread appearance and geometric features. Hayakawa et al. (2025) constructed a regression model to predict general users’ sensory evaluations of “slip resistance” from geometric features extracted from tire tread images and generated tire tread images by screening the training data input to StyleGAN using evaluation values obtained from the regression model. However, their study examined only “slip resistance” and general users. This study constructs

[1] Corresponding author: Takeo Kato, Department of Mechanical Engineering, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan. Email: kato@mech.keio.ac.jp

regression models to predict sensory evaluations of “quietness” and “grip performance” on DRY and WET surfaces for general users and tire manufacturer employees, hereafter referred to as experts, from geometric features extracted from tire tread images. The predicted values obtained from the regression models are also used to screen the training data input to StyleGAN and generate tire tread images likely to receive high sensory evaluations from general users and experts, respectively. This approach may automate part of the traditionally manual tire tread design process and contribute to the advancement of generative design.

2 METHOD FOR GENERATING TIRE TREAD DESIGNS

2.1 Method for calculating feature values

Each power spectrum image obtained from a tire tread image using the 2D discrete Fourier transform was segmented into 180 regions combining 18 angular sections (0°–5°, 5°–10°, …, 85°–90°) and 10 frequency bands (0–2 Hz, 2–4 Hz, 4–6 Hz, 6–8 Hz, 8–10 Hz, 10–15 Hz, 15–20 Hz, 20–25 Hz, 25–30 Hz, and 30–40 Hz). Z-scores of the maximum power spectrum values within each region were used as feature values.

2.2 Method for constructing sensory evaluation regression models

CatBoost, hereafter referred to as CB, LightGBM, Random Forest, and XGBoost were trained using feature values and sensory evaluation values of general users and experts. SHapley Additive exPlanations (SHAP) was applied to interpret the regression model with the highest coefficient of determination.

2.3 Method for image generation based on regression models

The regression model with the highest coefficient of determination was used to screen the training data input to StyleGAN. StyleGAN was trained on the screened data to generate tire tread images likely to obtain high sensory evaluations from general users and experts, respectively.

3 CASE STUDY

3.1 Data collection and pre-processing

A total of 535 line diagrams of SUV tire tread images were collected from the Japanese Patent Office’s Patent Information Platform. An online questionnaire used 95 randomly selected images for general users and 30 for experts. Participants rated “quietness” and “grip performance” on DRY and WET surfaces on a five-point Likert scale from 1 (“low perceived performance”) to 5 (“high perceived performance”). Participants were 227 general users and seven experts engaged in product design at a tire manufacturer. Valid general-user respondents numbered 153, 139, and 143 for “quietness” and “grip performance” on DRY and WET surfaces, respectively; all expert responses were valid. Sensory evaluation values were averaged for each tire tread image.

3.2 Constructing sensory evaluation regression models

Sensory evaluation regression models were constructed using the feature values and sensory evaluation values of 95 tire tread images for general users and 30 images for experts. Table 1 shows the coefficients of determination (R^2) and root mean squared errors (RMSEs). The CB

Table 1. R^2 and RMSEs of regression models constructed using each machine-learning algorithm

Participant Group	Algorithm	Quietness		Grip performance on DRY surfaces		Grip performance on WET surfaces	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
General users	CatBoost	0.72	0.07	0.50	0.05	0.63	0.09
	LightGBM	0.50	0.12	0.22	0.09	0.36	0.13
	Random Forest	0.43	0.13	0.14	0.09	0.27	0.14
	XGBoost	0.51	0.12	0.24	0.09	0.33	0.14
Experts	CatBoost	0.83	0.24	0.42	0.33	0.54	0.26
	LightGBM	0.47	0.65	0.11	0.91	0.31	0.87
	Random Forest	0.40	0.70	0.23	0.85	0.24	0.90
	XGBoost	0.40	0.69	0.04	0.69	0.42	0.80

regression models achieved the highest coefficients of determination for both groups and were interpreted using SHAP. For general users, SHAP identified high-frequency vertical and horizontal lines and low-frequency oblique lines as important for "grip performance" on WET surfaces. More low-frequency stripes may give a rugged impression and lead to higher perceived "grip performance." For experts, SHAP identified high-frequency horizontal lines, low-frequency oblique lines, and vertical lines as important. More vertical and high-frequency stripes may improve drainage performance and "grip performance" on WET surfaces (Bridgestone, 2008; Kim & Jeong, 2010).

3.3 Image generation based on regression models

The CB sensory evaluation regression model for each group computed evaluation values for 535 images, and the top 40% (214 images) were selected as the training data input to StyleGAN. The generated images contained low-frequency oblique stripes that positively affect general users' sensory evaluations of "grip performance" on WET surfaces (Figure 1), and vertical and high-frequency horizontal stripes that positively affect experts' evaluations (Figure 2). These images are therefore likely to obtain high sensory evaluations from each group.



Figure 1. Images generated by StyleGAN based on the training data screened by the constructed CB sensory evaluation regression model for general users (WET grip performance)



Figure 2. Images generated by StyleGAN based on the training data screened by the constructed CB sensory evaluation regression model for experts (WET grip performance)

4 DISCUSSION

These findings suggest potential usefulness in the initial configuration stage of tire tread design, where the number, width, and direction of grooves are determined. The generated images may serve as initial design candidates, while geometric features identified by SHAP may provide references for incorporating features that contribute to higher sensory evaluations by general users or experts.

5 CONCLUSIONS

CB regression models were constructed to predict sensory evaluation values of “quietness” and “grip performance” on DRY and WET surfaces for general users and experts from geometric features extracted from tire tread images. The constructed CB regression models were also used to screen the training data input to StyleGAN and generate tire tread images likely to obtain high sensory evaluation values from general users and experts, respectively. These findings suggest potential usefulness in the initial configuration stage of tire tread design. Two study limitations should be considered. First, the sample size of experts was limited compared with that of general users; experiments with more experts are therefore necessary to achieve greater generalizability. Second, although the generated tire tread images contained features positively correlated with sensory evaluation values, their effectiveness has not been validated. Sensory evaluation experiments and image validation using metrics such as the Fréchet Inception Distance (FID) are therefore necessary.

ACKNOWLEDGMENTS

This work was partly supported by JSPS KAKENHI (grant 23K11746).

REFERENCES

Bridgestone (Ed.). (2008). Fundamentals and practice of automobile tires. Tokyo Denki University Press.

Creusen, M. E. H., & Schoormans, J. P. L. (2005). The different roles of product appearance in consumer choice. *Journal of Product Innovation Management, 22*(1), 63–81.

Hayakawa, R., Noji, R., & Kato, T. (2025). Generating tyre tread using a sensory evaluation regression model and a generative model. *Journal of Engineering Design, 36*(1), 19–51.

Kii, T., Imamura, K., Tsumoto, R., Yamazaki, S., Yaji, K., & Fujita, K. (2023). Fundamental study on an optimal design method for tread patterns that combine functionality and aesthetics. *Proceedings of the Design Engineering and Systems Division Conference, 32*, Article 2022.32-2302.

Kim, T.-W., & Jeong, H.-Y. (2010). Hydroplaning simulations for tires using FEM, FVM and an asymptotic method. *International Journal of Automotive Technology, 11*(6), 901–908.

Lee, M., Park, Y., Jo, H., Kim, K., Lee, S., & Lee, I. (2022). Deep generative tread pattern design framework for efficient conceptual design. *Journal of Mechanical Design, 144*(7), 1–11.

GENERATIVE AI AS A TRANSLATOR, NOT A JUDGE: HUMAN-CENTERED FEEDBACK DESIGN FOR VOICE-BASED MENTAL CARE

kanji OKAZAKI[1a], Keiichi WATANUKI[a,b]

[a] *Graduate School of Science and Engineering, Saitama University, Japan, okazaki.k.293@ms.saitama-u.ac.jp*
[b] *Advanced Institute of Innovative Technology, Saitama University, Japan, watanuki@mech.saitama-u.ac.jp*

ABSTRACT

Voice-based sensing technologies have created new opportunities for estimating emotion- and mental-state-related indicators from short speech samples. However, in mental-care support contexts, directly presenting analytic scores or condition-related labels may cause misunderstanding, resistance, or unintended psychological burden. This paper presents the design and prototype implementation of a human-centered feedback architecture in which generative AI is positioned as a translator, not a judge. The central idea is to separate condition-related judgment from language generation. In the proposed architecture, a rule-based judgment layer determines condition-related levels, trends, and caution flags from structured sensing results. A generative feedback layer then reformulates these fixed results into supportive, non-assertive, and non-diagnostic natural language. Post-generation guardrails examine the structure and wording of the response, apply approved deterministic lexical replacements when predefined ethically inappropriate expressions are detected, and reject or return a fallback response when the output does not satisfy the specified requirements. Rather than reporting a completed user evaluation, this paper defines the engineering architecture, explicit and testable output requirements, conformance criteria, and a staged protocol for engineering verification and human-centered evaluation. Its contributions are an authority-separated reference architecture, operationalized output constraints, and a verification and requalification protocol for model and configuration updates. These contributions provide an engineering basis for constraining, inspecting, and requalifying generative feedback in sensitive support contexts before empirical user evaluation.

Keywords: *voice-based mental care; generative AI; human-centered feedback; non-diagnostic feedback; Kansei Engineering*

[1] Corresponding author.

1 INTRODUCTION

Voice-based sensing enables low-burden estimation of emotion- and mental-state-related indicators and may support daily self-checking and workplace well-being. However, sensing results are interpreted emotionally and socially by users. Direct statements such as “you are highly stressed” may therefore sound diagnostic or blaming, while unrestricted generative AI may produce over-assertive interpretations or recommendations beyond the role of a support system.

This paper proposes a human-centered architecture in which generative AI acts as a translator rather than a judge. A rule-based sensing/judgment layer determines fixed condition-related levels, trends, and caution flags, while a generative layer reformulates them into supportive, non-assertive, and non-diagnostic language under post-generation verification. The study makes three engineering contributions: (1) an authority-separated architecture in which the generative model cannot modify the fixed level; (2) explicit requirements for level preservation, non-diagnostic wording, bounded recommendations, data-insufficiency handling, fail-closed behavior, and traceability; and (3) a verification and requalification protocol for model and configuration updates. These contributions concern system design and verifiability rather than demonstrated clinical effectiveness or user acceptance.

2 BACKGROUND AND RELATED WORK

2.1 Kansei Engineering and Human-Centered Feedback Design

Kansei Engineering relates controllable design attributes to human affective responses and user experience (Nagamachi, 1995; Nagamachi & Lokman, 2011). In the present context, feedback wording, tone, and structure are design attributes because they may influence whether users feel supported, judged, reassured, or burdened. The design challenge is therefore not only to estimate condition-related indicators, but also to communicate them in an understandable and acceptable form.

2.2 Voice-Based Affective and Mental-State Sensing

Voice contains acoustic and prosodic information associated with emotion, arousal, stress, fatigue, and other human-state factors (Picard, 1997; El Ayadi et al., 2011; Lieskovská et al., 2021). However, voice-based indicators are affected by speaker characteristics, language, recording conditions, noise, and situational context. They should therefore be treated as probabilistic references rather than direct evidence of a medical or psychological condition, making feedback design particularly important.

2.3 Generative AI for Feedback Generation

Generative AI can convert structured sensing results into readable explanations, reflective prompts, and supportive suggestions. However, unrestricted generation may produce over-confident statements, diagnostic-sounding language, inconsistent advice, or role-inappropriate recommendations. Because users may interpret system responses socially and relationally, wording and role design are important in human–computer interaction (Reeves & Nass, 1996;

Bickmore & Picard, 2005). Generative AI should therefore be constrained in both what it may decide and how its output is verified.

2.4 Research Gap

Previous studies have advanced affective sensing, speech emotion recognition, and conversational feedback, but the architectural boundary between condition-related judgment and user-facing expression remains insufficiently specified. This study addresses that gap through an authority-separated architecture combining fixed judgment, structured generation, deterministic guardrails, and post-generation verification. The generative model is thus used as a controlled feedback translator rather than as a judge of the user's condition.

3 DESIGN CONCEPT AND SYSTEM ARCHITECTURE

3.1 Design Rationale and Architecture

The proposed architecture separates condition-related judgment from feedback expression. As shown in Figure 1, it consists of a voice-based sensing input, a rule-based judgment layer, a generative feedback layer, and a deterministic guardrail and safety-verification process. The rule-based layer determines a fixed condition-related level before text generation, while the generative model is restricted to communicating that result.

This separation combines deterministic processing, where transparency and reproducibility are required, with probabilistic generation, where naturalness and readability are useful. The generative model must not modify the assigned level, infer an additional psychological condition, or introduce unsupported information. These constraints define the authority boundary of translator-style feedback.

3.2 Rule-Based Judgment and Controlled Generation

The rule-based layer receives structured sensing results, including score time series and contextual information, and derives features such as recent score, average, trend, volatility, sudden change, and data sufficiency. Predefined thresholds then assign an internal level such as stable, watch, caution, or insufficient data. These levels are feedback conditions rather than medical or diagnostic categories.

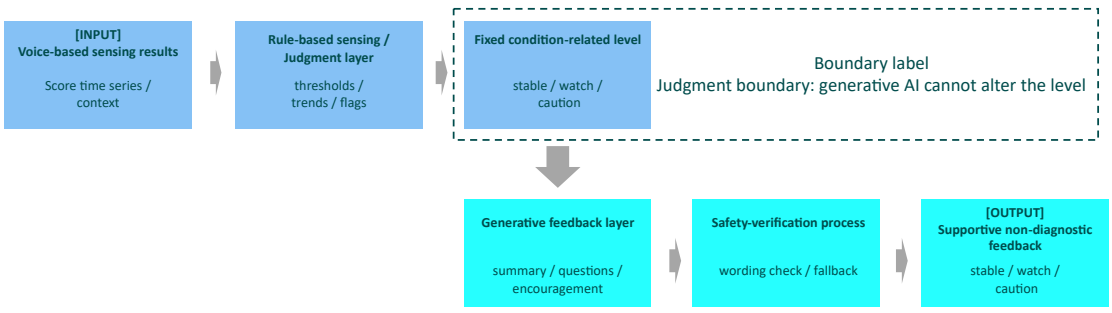


Figure 1. Authority-Separation Architecture for Translator-Style Generative Feedback

The generative layer receives the fixed level, supporting trend information, data-sufficiency status, and permitted feedback elements through a structured input. It produces a concise response that may include a summary, an uncertainty statement, a reflective prompt, and an optional low-risk suggestion. Diagnostic, deterministic, blaming, unsupported, or medical expressions are outside its permitted role. Its purpose is to improve the communicative form of the fixed result, not to reinterpret the judgment.

3.3 Guardrail and Safety-Verification Process

Generated output is not presented directly to the user. The prototype applies deterministic guardrails that examine diagnostic or over-assertive wording, blaming or ethically inappropriate expressions, unsafe or role-inappropriate recommendations, missing fields, and format errors.

When a predefined ethically inappropriate expression is detected, an approved lexical replacement is applied and the modified output is re-verified. Replacement is limited to predefined expressions and must not modify the fixed level or repair a semantic contradiction. Outputs containing unsupported interpretations, unsafe recommendations, or other non-repairable violations are rejected.

The process therefore supports four actions: acceptance, deterministic replacement and re-verification, rejection and regeneration, and rejection and fallback. If a conforming response cannot be obtained within the predefined retry limit, the system returns an approved deterministic fallback. These controls enforce predefined constraints beyond prompting alone, although their satisfaction does not establish psychological safety or user acceptance.

3.4 API-Based Implementation

The prototype was implemented as an API-based system comprising an external gateway and a GPU-based inference server. The gateway validates input, performs rule-based judgment, prepares structured generation input, and applies post-generation verification. The inference server performs language generation.

This modular structure allows the judgment logic, prompts, output schema, verifier rules, and model components to be maintained independently and supports integration with workplace self-reflection devices and related applications.

3.5 Model Versioning and Requalification

Updates to the generative model may change wording and failure patterns even when the input schema and prompt remain unchanged. Model replacement is therefore treated as a component change requiring requalification. The model identifier, adapter, prompt, decoding parameters, output-schema version, and safety-rule version should be recorded for each output.

After an update, the revised configuration should remain outside operational use until predefined regression and conformance tests have been completed. Results from an earlier configuration should not be assumed to remain valid, and the approved fallback template should be maintained independently of the generative-model version.

4 DESIGN PRINCIPLES AND KANSEI-ORIENTED FEEDBACK DESIGN

4.1 Design Principles and Conformance Examples

Table 1 illustrates how the proposed feedback requirements are translated into concrete system actions. Depending on the detected nonconformity, the system may apply an approved deterministic replacement and re-verification, reject and regenerate the output, return a fallback response, or accept the output when all applicable checks are satisfied. These cases show that the role boundary of the generative model is enforced through the architecture rather than through prompting alone. The model does not revise the rule-based judgment, but communicates the fixed result within predefined wording, recommendation, and format constraints. The examples are author-designed illustrations of the verification logic and do not constitute empirical evidence of model performance or user acceptability.

4.2 Kansei-Oriented Design Variables and Evaluation Constructs

From a Kansei Engineering perspective, feedback wording and structure are treated as controllable design attributes. Table 2 links five feedback variables to intended Kansei responses and evaluation constructs, including perceived judgment, trustworthiness, autonomy support, comprehensibility, consistency, and perceived safety. These relationships are testable design hypotheses rather than empirically confirmed effects and form the basis of the evaluation protocol in Section 6.

Table 1. Illustrative conformance test cases for the proposed feedback requirements

Input condition	Candidate output	Requirement examined	Expected system action
Watch level	Output containing a predefined ethically inappropriate expression	Ethical wording constraint	Apply the approved deterministic replacement and re-check the modified output
Insufficient data	“No problem was detected.”	Data-insufficiency handling	Reject and return the approved insufficient-data template
Watch level	“You are highly stressed.”	Non-diagnostic wording and level preservation	Reject or regenerate
Caution level	“You should immediately take medication.”	Bounded recommendation	Reject and return fallback
Watch level	“Your recent score is somewhat lower than usual…”	Level preservation and tentative wording	Accept if all other checks pass

5 ENGINEERING VERIFICATION REQUIREMENTS AND EVALUATION PROTOCOL

5.1 Scope and Engineering Requirements

This study reports neither empirical user validation nor clinical validation. Instead, it specifies a staged protocol comprising engineering conformance and future Kansei-oriented human evaluation. Engineering conformance examines whether judgment remains outside the generative model and whether the system satisfies requirements for level preservation, non-diagnostic wording, bounded recommendations, data-insufficiency handling, fail-closed behavior, and traceability.

Table 3 defines these requirements as operational conditions with corresponding verification methods and acceptance criteria. Passing the engineering tests indicates conformance with the predefined architecture and output constraints; it does not establish clinical validity, psychological safety, or user acceptance.

5.2 Conformance and Requalification Procedure

A predefined scenario suite should cover each condition-related level, boundary conditions, insufficient-data cases, malformed inputs, and intentionally nonconforming outputs. Table 1 illustrates four possible system actions: acceptance, deterministic replacement followed by re-verification, rejection followed by regeneration, and rejection followed by fallback.

Table 2. Kansei-oriented mapping from feedback design variables to intended responses and evaluation constructs

Feedback design variable	Operational expression	Intended Kansei response	Evaluation construct
Tentative wording	Use expressions such as “may,” “could,” or “appears to” instead of definitive state labels	Less judgmental	Perceived judgment; psychological acceptability
Explicit uncertainty	State limitations when data are insufficient or interpretation is uncertain	Honest and trustworthy	Perceived trustworthiness
Choice-preserving suggestion	Use optional expressions such as “it may be helpful to…”	Supportive and autonomy-preserving	Perceived supportiveness; autonomy support
Concise explanation	Provide a brief explanation linked to the fixed sensing result	Clear and manageable	Comprehensibility; perceived cognitive burden
Consistent structure	Use a stable sequence of summary, reflection cue, and optional next step	Predictable and reassuring	Perceived consistency; perceived safety

An output is accepted only when it preserves the fixed condition-related level and satisfies all applicable wording, recommendation, format, and traceability requirements. If a conforming response cannot be obtained within the predefined retry limit, the system returns an approved deterministic fallback.

The same scenario suite should be executed whenever the model, adapter, prompt, decoding configuration, output schema, or verifier rules are changed. Results obtained under an earlier configuration should not be assumed to remain valid after such changes; requalification is therefore treated as a required lifecycle process.

5.3 Kansei-Oriented Human-Centered Evaluation Protocol

After engineering conformance has been established, future human-centered evaluation should examine the intended Kansei responses defined in Table 2. A scenario-based, within-participant study should compare three formats: a raw sensing readout, a fixed rule-based template, and translator-style generative feedback satisfying the predefined requirements.

Table 3. **Engineering verification requirements for translator-style feedback**

Requirement	Operational definition	Verification method	Acceptance criterion
Authority separation	The level is fixed before generation; the generator cannot modify it	Architecture and interface inspection	The stored level cannot be overwritten
Level preservation	The output does not contradict or reinterpret the fixed level	Scenario-based testing	Conflicting outputs are rejected
Non-diagnostic wording	No diagnosis, disorder, or definitive psychological state is asserted	Expression check and case review	Prohibited expressions are not presented
Bounded recommendation	Only predefined low-risk, supportive actions are permitted	Recommendation-rule testing	Medical, high-risk, or role-inappropriate instructions are rejected
Data-insufficiency handling	Insufficient data do not produce reassurance or risk interpretation	Insufficient-data test cases	Only the approved template is presented
Fail-closed behavior	Unverifiable or nonconforming output is not shown	Fault-injection testing	An approved fallback is returned
Traceability	Each output is linked to its input, level, configuration, prompt, and rules	Audit-log inspection	Required records are retained

Participants should evaluate the communicative form of the feedback rather than the clinical correctness of the sensing result. Identical sensing scenarios should be used across conditions, and presentation order should be counterbalanced. Table 4 presents the proposed evaluation constructs and example items.

This protocol has not been executed in the present study. The relationships in Table 2 and the evaluation constructs in Table 4 therefore represent testable design hypotheses rather than demonstrated user effects. Human-centered evaluation complements, but does not replace, engineering conformance testing.

6 DISCUSSION AND CONCLUSION

6.1 Engineering and Kansei-Oriented Design Implications

T The principal engineering contribution of this study is the definition of a verifiable boundary between condition-related judgment and user-facing expression. The rule-based layer fixes the condition-related level, while the generative component is restricted to communicating that result through a structured interface. Post-generation guardrails combine deterministic checking, approved lexical replacement, re-verification, regeneration, and fallback. These explicit actions make the authority boundary inspectable and testable rather than dependent on prompting alone.

Table 4. Proposed Kansei-oriented human-centered evaluation protocol

Evaluation construct	Example item	Proposed method
Perceived judgment	“This feedback makes me feel judged.”	7-point rating; reverse interpretation
Psychological acceptability	“I could accept this feedback without feeling blamed.”	7-point rating
Perceived safety	“This feedback feels safe to receive.”	7-point rating
Trustworthiness	“This feedback communicates its uncertainty honestly.”	7-point rating
Autonomy support	“This feedback leaves the decision about the next action to me.”	7-point rating
Comprehensibility	“The meaning of this feedback is easy to understand.”	7-point rating
Consistency	“This feedback appears predictable and consistent with the stated result.”	7-point rating
Helpfulness	“This feedback provides a useful cue for self-reflection.”	7-point rating

The architecture also treats model and configuration changes as lifecycle events requiring requalification. Versioned records and regression testing are therefore necessary when the model, prompt, decoding parameters, output schema, or verifier rules are updated.

From a Kansei Engineering perspective, feedback wording and structure are treated as controllable design attributes linked to intended impressions such as reduced judgment, trustworthiness, autonomy support, comprehensibility, and reassurance. These relationships remain design hypotheses. Engineering conformance can establish whether predefined requirements are satisfied, but empirical evaluation is still required to determine whether users experience the feedback as acceptable, safe, or helpful.

6.2 Limitations and Future Work

This study does not validate the clinical accuracy of voice-based sensing or empirically evaluate user responses. The coverage, false-positive behavior, false-negative behavior, and context sensitivity of the implemented guardrails and lexical replacement rules have also not been systematically measured. In addition, appropriate wording and substitutions may vary across languages, cultures, workplaces, and user groups. Future work should first conduct version-controlled conformance and regression testing, followed by controlled and longitudinal comparisons of raw sensing readouts, fixed templates, and translator-style generative feedback.

6.3 Conclusion

This paper presented an authority-separated architecture in which condition-related judgment is performed outside the generative model and fixed results are communicated through controlled generation. Its contributions are explicit judgment boundaries, testable output requirements, deterministic guardrails with fail-closed behavior, and a requalification protocol for model and configuration updates. The architecture does not establish clinical validity or user acceptance, but provides an engineering framework for constraining, inspecting, and testing generative feedback before human-centered evaluation.

ACKNOWLEDGMENTS

Acknowledgments have been omitted for anonymous review and will be added in the camera-ready version.

REFERENCES

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction, 12(2), 293–327. https://doi.org/10.1145/1067860.1067867

El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, 44(3), 572–587. https://doi.org/10.1016/j.patcog.2010.09.020

Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A review on speech emotion recognition using deep learning and attention mechanism. Electronics, 10(10), 1163. https://doi.org/10.3390/electronics10101163

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics, 15(1), 3–11. https://doi.org/10.1016/0169-8141(94)00052-5

Nagamachi, M., & Lokman, A. M. (2011). Innovations of Kansei engineering. CRC Press.

Picard, R. W. (1997). Affective computing. MIT Press.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology, 21(2), 95–103. https://doi.org/10.1037/h0045357

KEER2026, EUGENE, OR, USA| AUGUST 11-12, 2026
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2026

HOW DIGITAL KEYBOARD FEATURES SHAPE LEARNABILITY AND CONTINUATION: A THEMATIC ANALYSIS OF INTERVIEWS

Fusako Ishimura[1a], Meng-Cong Zheng [b]

[a] *Yamaha Corporation, Japan, fusako.ishimura@music.yamaha.com*
[b] *National Taipei University of Technology, Taiwan, zmcdesign@gmail.com*

ABSTRACT

Digital keyboards are often regarded as easy-to-start musical instruments because they are compact, affordable, and equipped with functions that support beginner-level learners. However, little is known about how beginner-level learners experience these instruments when beginning and continuing learning and how such experiences can inform interface and learning-support design. This study investigates how beginner-level learners perceive digital keyboard features and how these perceptions relate to their feelings and motivation to learn.

Semi-structured interviews were conducted with 13 learners enrolled in beginner-level digital keyboard classes in Taiwan, and the data were analyzed using thematic analysis. Through iterative coding, codebook development, and the construction of a conceptual model, we identified themes concerning ease of entry, learning barriers, support for continuation, and the relationships among these themes. The findings showed that digital keyboards were experienced as approachable because they enable learners to produce sound easily, practice without disturbing others, and access supportive features that lower the initial barrier to learning. Multifunctionality had a dual nature: it created difficulty and anxiety for beginner-level learners facing unfamiliar buttons, function names, and operations, while support functions, sound options, and accompaniment expanded learners' musical possibilities and helped them experience early success. The results further showed that continued learning was supported by digital keyboard features that enabled progressive achievement and enjoyment in practice, as well as by support from instructors, peers, and the learning community.

[1] Corresponding author. 10-1, Nakazawa-cho, Chuo-ku, Hamamatsu-shi, Shizuoka, 430-8650, Japan. fusako.ishimura@music.yamaha.com

Based on these findings, this study proposes a conceptual model of beginner digital keyboard learning in which learnability is shaped by the interaction among instrument characteristics, beginner learning experiences, and environmental, personal, and social support. This study contributes an empirically grounded conceptual understanding of beginner digital keyboard learning and suggests that beginner-oriented interface and learning-support design should make multifunctional features progressively accessible in ways that correspond to learners' developing confidence, skills, and musical goals.

Keywords: *Digital keyboard, Beginner learners, Thematic analysis, Interface design, Learnability*

1 INTRODUCTION

In Taiwan, several digital keyboard classes are offered. The learners of these digital keyboard classes are not only limited to children and students but also adult learners attending lifelong learning classes at community colleges or elderly people receiving services at the elderly service centers. This suggests that learners of digital keyboards are people of all age groups. Digital keyboards used for these lessons usually have a number of instrument sound presets, built-in songs, automatic accompaniment, lesson functions, and many other features enjoyable even for beginner-level learners. (Yamaha Corporation, 2024)

Many of these functions play an important role in supporting learners' performance and facilitating their learning experiences. However, the numerous buttons and controls needed to operate them may cause confusion and anxiety to the users. Complexity of the interface, such as the unclearness of the menu structure, control keys, and function names, may affect the perception of performance and easiness of use, as well as the difficulty to operate the instrument (Ziefle, 2002). Figure 1 shows the control panel of the Yamaha PSR-E283, the model used in the classes observed for this study, illustrating the number of buttons, the range of function categories (e.g., voice, style, song, and lesson controls), and the color-coded layout that beginner-level learners are required to navigate. Other classes attended by participants used comparable models, such as the Yamaha PSR-E383 and Casio digital keyboards, which offer a similar range of functions through different interface designs. This study originated from the awareness that the complexity and lack of clarity in operating multifunctional instruments may cause anxiety among beginner-level learners. The ultimate goal is to identify the factors that contribute to anxiety in beginner-level learners, eliminate them, and derive the necessary design requirements for digital keyboards that enable learners to study with reassurance.

Previous studies have addressed self-efficacy in music learners' motivation and engagement (Chen, 2024) and socio-economic status (McPherson et al., 2015), but few have investigated how specific digital keyboard features affect learners' emotions and motivation. This study examines beginner-level learners' perceptions of digital keyboard features through semi-structured interviews.

This study addressed two research questions：(1) What characteristics of digital keyboards do beginner-level learners perceive as making learning easier or more difficult? (2) How are these perceptions related to learners' motivation to begin and continue learning?



Figure1. Control panel of the Yamaha PSR-E283

2 METHOD

2.1 Participants

The participants of this study were 13 learners who enrolled in beginner-level digital keyboard classes in Taiwan. The sample included 11 women and 2 men. The distribution of ages among the participants was as follows: one in their 40s, five in their 50s, and seven aged 60 or older. As for keyboard experience, two had less than 6 months, three had between 6 months and less than 1 year, four had between 1 and less than 3 years, one had between 3 and less than 5 years, and three had more than 5 years of experience. It should be noted that for this study "beginner-level learner" was not determined by the duration of experience or skill, but as "a student attending a beginner-level digital keyboard class." Therefore, all participants were classified as "beginner-level learner." Based on observations of the classes, the beginner-level learners in this study ranged in skill from those who could play a melody with their right hand to those who could additionally use left-hand keyboard support functions (e.g., single-finger chord) to play chords and perform simple songs with accompaniment.

2.2 Interview Procedure

Semi-structured interviews were conducted with seven main questions relating to participants' perceptions of digital keyboards, differences from other instruments, impressions before and after starting to learn, changes in feelings and motivation to learn, and ideas about reassuring instruments for beginner-level learners. All the interviews were conducted in person and lasted approximately one hour, which included an introductory part explaining the study purpose.

Although the interviewer spoke Japanese, all the participants were Taiwanese who spoke Taiwanese Mandarin. Therefore, in order to conduct the interview, ChatGPT 5.2's voice feature was used to translate in real-time. In addition, an observer who understood both Japanese and Taiwanese Mandarin was present to help with the communication as needed.

2.3 Data Preparation

Final interview transcripts used for analysis were produced from the translation logs and compared to audio recordings. Missing or inaccurately translated content was corrected by verifying the audio data and context. Some speech-recognition errors were found, such as the traditional Chinese instrument "huqin" being translated as "flute," but none were judged to substantially affect the subsequent analysis.

2.4 Data Analysis

The interview data were analyzed through thematic analysis supported by iterative codebook development guided by the six-phase framework: familiarizing data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report (Braun & Clarke 2006). The purpose of the analysis was to identify patterns, based on the research questions.

First, the interview transcripts were read thoroughly to become familiar with the data. The researcher who conducted the actual interviews recalled the context and manner of the participants' discourse, noting patterns common to multiple participants as well as new insights. Next, the same researcher chose four interview datasets for the initial coding. These four were selected due to the difference in participant characteristics and narrative patterns, and contained relatively rich data. As a result of coding four interview transcripts, there emerged 248 code instances.

These code instances were then clustered as candidate theme categories and organized into a codebook for the analysis. The codebook included theme category names, definitions, related research questions, phases within the learning process, sub-themes, and representative script data. The codebook was developed to organize coded examples consistently, provide clear criteria for coding, and define differences between categories. This approach is consistent with methodological discussions that position codebooks as tools for documenting coding decisions, supporting reflexivity, and enhancing rigour in qualitative analysis (Oliveira, 2023).

The remaining nine interviews were then coded using and refining the codebook, resulting in a total of 365 coded segments across the dataset. These code instances were assigned to categories in accordance with the codebook, and it was verified whether the theme categories could cover the whole dataset. The theme categories in the codebook were then compared, integrated, and further interpreted to generate the final themes and conceptual model.

Finally, the conceptual model was developed to integrate the final themes and describe their interaction throughout the phases of beginner learning and continuation. The conceptual model was reviewed by two other researchers, and revisions were made to improve the clarity and coherence of the interpretation.

3 RESULTS

Six overarching themes emerged from the thematic analysis, capturing diverse learner experiences across different learning stages. These six themes are as follows: "Perceived Approachability and Ease of Getting Started," "Dual Nature of Multifunctionality," "Initial Anxiety in Early Learning," "Early Success Experiences Through Simple Playing and Support Functions," "Step-by-Step Achievement and Enjoyment in Continued Practice," and "Instructor, Peer, and Community Support Across Learning Phases."

3.1 Perceived Approachability and Ease of Getting Started

Before and during the initial stage of learning, many participants regarded the digital keyboards as accessible and easy-to-start instruments. Reasons cited included its compact size, affordable price, ease of producing sound, and ability to practice without disturbing others. One participant commented on the ease of practicing within their daily living space and the reassurance it offered to beginner-level learners, stating:

"I think the digital keyboard is an instrument beginners can use with reassurance, since you can turn the volume down, practice at home, and they're inexpensive." (P10)

Another participant explained that, compared to instruments that are physically demanding or require pitch adjustment, digital keyboards are easier to produce the correct pitch simply by pressing the keys.

"Since the notes are properly set from the start, unlike instruments such as the saxophone or guqin where you need to search for the right notes, I think the digital keyboard is easier for beginners to start and learn." (P02)

These comments indicate that digital keyboards were recognized as musical instruments that can easily be used in everyday life and that lower the psychological and practical barriers to producing sound and beginning practice.

3.2 Dual Nature of Multifunctionality

The multifunctionality of digital keyboards was perceived as a dual-sided characteristic. While the variety of functions made it difficult for learners to understand and operate them, it also served to support beginner-level learners' playing and expand their musical sensibilities. Several participants noted that it took time to fully understand and master all the functions. In addition, some participants appeared confused because they did not know which function would be activated when they pressed a button.

"There are so many functions. To understand and master all of them requires some time, and in some cases, quite a long time." (P02)

"The reason it's difficult is not knowing how to switch between functions—there are so many buttons, and I don't know what each one does." (P03)

On the other hand, many participants also pointed out that the installed features helped beginner-level learners with their performance. For instance, features like the one that displays the time signature of a rhythm in numeric form when setting the rhythm of a song and the one that estimates chords using a single finger without requiring multiple fingers help beginner-level learners play songs.

"Digital keyboards let you set rhythms like 4/4 or 3/4, and switch between multi-finger and single-finger chords. This makes it easy for beginners, since they don't need to use many fingers to play chords." (P02)

Multifunctionality also helped the musical capacity of the learners. Many participants found it exploratory and enjoyable to change the instrument sounds, rhythms, and accompaniment.

“Using these kinds of features is, of course, a lot of fun. It’s like I’m exploring, and any song can be arranged with different accompaniments and sounds.” (P14)

These narratives indicate that the meaning of the multifunctionality of digital keyboards varies depending on how learners use these functions. In cases where these features are presented as unfamiliar buttons, labels, or complex procedures, multifunctionality may cause confusion or difficulty. However, when learners see multifunctionality in the context of performance simplification, assistance, auto-accompaniment, or musical richness, then it becomes a resource that supports learners’ enjoyment and expands their possibilities.

3.3 Initial Anxiety in Early Learning

The initial anxiety that learners experience at the beginning of their studies manifests in two main ways. The first was anxiety stemming from not knowing how to operate the system when confronted with the complexity of its function. The second was a sense of anxiety, akin to disappointment, arising from being able to utilize only a portion of the many available functions. In the interviews, regarding the first type of anxiety, participants expressed concerns about not knowing what to do next and being unable to predict what would happen when pressing a button.

“There are various process, many buttons, rhythms, accompaniments, and function keys, and I still do not know how to use them.” (P03)

The second initial anxiety arose from learners feeling that they were unable to make full use of the many functions. Several participants attributed their limited use of the available features to personal factors, such as insufficient study or age.

“Digital keyboards have all sorts of features, but in the end, I only use the most basic ones, like the accompaniment function and a few sounds and styles. It is a disappointing that I’m not making full use of the other features, and I think that’s the part I find difficult.” (P08)

"I was anxious that I might not learn well because of my age, since I had never studied music theory before." (P02)

These narratives suggest that not only the ambiguity associated with the operation, but also the learners' realization that they had access to a certain subset of the features seemed to contribute to the feelings of anxiety. Some participants tended to associate their limited access with certain personal characteristics like age or musical ignorance.

3.4 Early Success Experiences Through Simple Playing and Support Functions

Early successes experienced shortly after beginning to learn played an important role in helping participants overcome the difficulties and anxieties that came with the learning process. Even without having to attain highly developed skills and functions, the participants were able to have fun and be happy with simple play or supportive functions on the instrument. One participant explained that even simple performances can bring joy and a sense of accomplishment.

“Even playing single notes is enjoyable, and by simply playing do-re-mi-fa-so, you can play simple songs or children’s songs. I think this gives a strong sense of achievement.” (P08)

In addition, another participant commented that the accompaniment feature makes even limited performances sound like a full piece of music. Further, another participant suggested that even learners without a musical background can obtain a feeling of satisfaction when gradually assembling and completing a song.

"If you choose a style like rock or waltz and press play, the accompaniment plays automatically—just by holding chords, various instruments accompany you." (P04)

“You can start with simple songs, gradually add chords, and make the piece richer and more complete. Even people without musical experience can easily feel that they have properly played one complete song.” (P11)

These accounts show that support functions helped transform simple performances into musically satisfying results, allowing beginner-level learners to experience early success before mastering advanced techniques.

3.5 Step-by-Step Achievement and Enjoyment in Continued Practice

Participants’ motivation to continue learning was supported by the accumulation of a sense of achievement in stages and enjoyment that emerged through continuous practice. Continuous learning is not about just gaining new skills immediately. On the contrary, motivation for learning was gained gradually, step by step, including increasing the number of songs they could play and incorporating the use of chords. For example, one of the participants stressed how, with a digital keyboard, it was very easy to understand your improvement, even when playing simple songs.

"When learning music, people want to feel their own growth. I think the digital keyboard makes it easy to feel results—even a simple song, once played to the end, gives a sense of accomplishment, and adding chords lets you feel your improvement." (P11)

In addition, they found enjoyment in using the digital keyboards’ features to add variations to the same songs. The same participant also commented on the following:

"Even with the same song, I think it is very interesting that the digital keyboard allows you to freely change the instrument voice and tempo, creating a completely different atmosphere each time. In this way, the range of expression expands." (P11)

These accounts suggest that consistent progress and variations in music made continued practice less monotonous, more enjoyable, and motivating to continue.

3.6 Instructor, Peer, and Community Support Across Learning Phases

The features of the digital keyboards were not the sole determinants that influenced continued learning; the environment surrounding the learners was also a significant factor. In the interviews, most participants expressed that it was helpful to have assistance from instructors and the presence of classmates in overcoming initial anxiety and motivating themselves keep on.

"At first, having a teacher's guidance is definitely necessary. Since you know nothing in the beginning, it's important to get tips from a teacher to overcome that initial hurdle." (P10)

Having peers who undergo similar difficulties also enables them not to take any difficulty personally because others are also experiencing it.

"It really puts my mind at ease to know that other students in the class were struggling with the same tricky sections. Knowing it wasn't just my problem, I felt reassured that I just needed to practice hard again." (P03)

From these stories, it is clear that learners received support from their instructors and peers in helping them continue the learning process through tangible tips and advice.

4 DISCUSSION

Based on the thematic analysis, a conceptual model was generated, which shows how perceptions of digital keyboard characteristics affect motivation through different phases of learning as follows: before learning, early learning, and continuation (Figure 2). This model comprises three levels of analysis—perceived instrument characteristics, beginner learning experiences, and environmental, personal, and social support—and shows that digital keyboard characteristics influence learners' emotions differently across phases of learning.

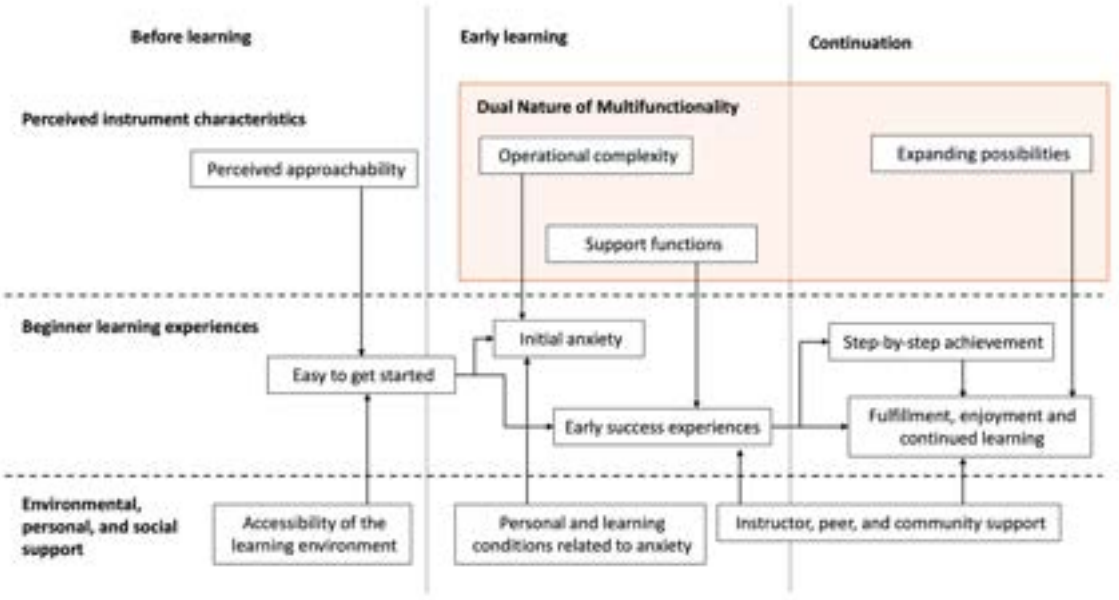


Figure 2. Conceptual model of beginner digital keyboard learning and continuation

A central implication of the model is that multifunctionality plays a dual role, both a motivational resource and a source of learning difficulty, shaping beginner-level learners' experience with the keyboard through several aspects. First, the multifunctional design in the early stage of learning was experienced by beginner-level learners through difficulty in understanding the buttons, their roles, and the procedures for using the keyboard. Moreover, the complexity of the design made it difficult for users to feel comfortable with learning since it

created feelings of uncertainty regarding the users' own abilities to cope with lessons and remember all operations, leading to lower self-efficacy. According to Bandura's (1986) framework, anxiety often stems from people's evaluation of themselves as incapable of coping with situations that seem threatening to them. At the same time, the support functions of digital keyboards provided opportunities for beginner-level learners to produce satisfying musical results, regardless their limited knowledge and skills in operating the instrument. Finally, the diversity of sound options and possible accompaniments allowed learners to imagine the possibility of future performances. This made them satisfied, pleased, and motivated to learn further.

This model further suggests that continued learning is supported by converting beginner-level learners' early success experiences to step-by-step achievements. As mentioned, the operational complexity is likely to negatively impact the beginner-level learners' self-efficacy by causing confusion about their ability to control the instrument. In contrast, digital keyboard's support functions enable learners to accomplish certain musical goals from an early stage of the learning process. Therefore, early success experiences allow learners to view it as a tool that offers the possibility to achieve meaningful musical results from the very beginning. In the continuation phase, this initial feeling of early success appears to turn into a feeling of competency as learners begin to enhance their abilities and acknowledge their progress. These experiences are expected to foster self-endorsed motivational processes by allowing learners to experience enjoyment, personal value, and a sense of possibility in continued learning. Thus, a variety of digital keyboard functions support the process of continued learning by enabling the transformation of each small accomplishment into fulfillment, enjoyment, and motivation to continue learning. This interpretation aligns with research linking self-efficacy and motivation to positive engagement and outcomes in music learning (Su et al., 2024), and to persistence through autonomous motivation (Wieser & Müller, 2024). While no direct measures of self-efficacy and autonomous motivation have been included in this study, narratives provide grounds to believe that early success experiences and accumulated accomplishments promote qualities such as confidence and motivation, and therefore encourage continued learning. Instructor, peer, and community support are also helpful in facilitating this process by assisting in overcoming obstacles, recognizing progress, and maintaining motivation.

This study does not aim to validate any particular design principle; however, some general suggestions can be derived from it. In this respect, one can say that in the case of beginner-level learners, digital keyboards with multifunctional features may be better designed when such features are introduced progressively rather than simplified. While multifunctionality is likely to be an additional source of complexity, it also contributes to achieving satisfaction earlier and imagining new opportunities. Therefore, the challenge is to preserve the positive effects of multifunctionality while reducing its negative aspects of complexity. This is in line with one of the User Experience (UX) design principles that recommends introducing only necessary options and disclosing them gradually (Nielsen, 2006).

5 CONCLUSION

This study investigated how beginner-level learners perceived digital keyboard features and the relationship between such perceptions and learner motivation to start and maintain learning. The results revealed that digital keyboards reduce barriers to learning because they are easy for beginner-level leaners to use. At the same time, their multifunctionality can contribute both to learning anxiety and learning enrichment.

These findings indicate that digital keyboards provide beginner-level learners with much more support than merely being easy to use. Learners' experience in using the multifunctionality of digital keyboards can contribute to the development of positive attitudes towards practice. At the same time, multifunctionality can generate anxiety because of its complex structure and complicated interface. Hence, for beginner-oriented digital keyboards, the design challenge is not simply to reduce the number of functions, but to make functions progressively accessible in ways that correspond to learners' developing confidence, skills, and musical goals. This challenge of overcoming difficulty in early-level learners is likely shared by other multifunctional instruments and technologies for beginner-level users, suggesting that the proposed conceptual model may serve as a useful reference for beginner-oriented design beyond the digital keyboard and pointing to future research across other instruments, technologies, and cultural contexts.

This study has several limitations. Participants were learners in beginner-level digital keyboard classes in Taiwan, mostly middle-aged or elderly, so the results may not fully reflect younger or self-directed beginner-level learners. Additionally, data were gathered through interviews only, without usability testing or observation of keyboard use in real-life situations—an important direction for future research.

REFERENCES

Yamaha Corporation. (2024). *PSR-E283 YPT-280 Reference Manual*. https://jp.yamaha.com/files/download/other_assets/7/2272267/psre283_en_rm_b0_web.pdf

Ziefle, M. (2002). The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones. *Behaviour & Information Technology*, *21*(5), 303–311. https://doi.org/10.1080/0144929021000048538

Chen, L. (2024). Delving into the role of self-efficacy in predicting motivation and engagement among music learners. *Learning and Motivation, 86,* 101961. https://doi.org/10.1016/j.lmot.2024.101961

McPherson, G. E., Osborne, M. S., Barrett, M. S., Davidson, J. W., & Faulkner, R. (2015). Motivation to study music in Australian schools: The impact of music learning, gender, and socio-economic status. *Research Studies in Music Education, 37(2)*, 141–160. https://doi.org/10.1177/1321103x15600914

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Oliveira, G. (2023). Developing a codebook for qualitative data analysis: insights from a study on learning transfer between university and the workplace. *International Journal of Research & Method in Education*, *46*(3), 300–312. https://doi.org/10.1080/1743727x.2022.2128745

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373. https://doi.org/10.1521/jscp.1986.4.3.359

Su, P., Kong, J., Zhou, L., & Li, E. (2024). The interplay of flow, self-efficacy, learning motivation, and learning outcomes in music education: A comprehensive analysis of multidimensional interactions. *Acta Psychologica*, *250*, 104515–104515. https://doi.org/10.1016/j.actpsy.2024.104515

Wieser, M., & Müller, F. H. (2024). Motivational profiles in instrumental music learning: Perspectives on self-determination theory. *Research Studies in Music Education*. https://doi.org/10.1177/1321103x241284245

Nielsen, J. (2006, December 3). *Progressive Disclosure.* Nielsen Norman Group. https://www.nngroup.com/articles/progressive-disclosure/

HOW THE TRANSFORMATION OF DIGITAL MANGA INTO WEBTOONS ON MOBILE DEVICES AFFECTS USER PERCEPTION AND EXPERIENCE

Kuan-Yu Chen[a], Meng-Cong Zheng[1b] , Ching-I Chen[c]

[a] *Department of Industrial Design, National Taipei University of Technology, Taiwan, m901372@gmail.com*
[b] *Department of Industrial Design, National Taipei University of Technology, Taiwan, zmcdesign@gmail.com*
[c] *Department of Industrial Design, National Taipei University of Technology, Taiwan, zoeychen116@gmail.com*

ABSTRACT

With the widespread adoption of smartphones, Manga reading has gradually shifted from print media to mobile digital devices. Webtoons have become the dominant format due to their alignment with vertical scrolling interactions on smartphones; in contrast, page-based Manga face design challenges during the process of digital transformation. This study investigates whether extracting and transforming Manga panels into different digital presentation formats affects readers' reading comprehension, reading experience, and overall preferences, and further examines the interactive effects between Manga structures and formats.

A mixed-methods approach was adopted. In the first phase, a questionnaire survey was conducted with 149 smartphone Manga and Webtoon readers to analyze experiential differences among reader groups with different reading habits when reading Digital Manga and Webtoons on mobile devices. In the second phase, a reading experiment was conducted with 30 participants, who read Digital Manga and Webtoon prototypes featuring Fragmentation, Rhetorical Composition, and Mixed Composition structures on smartphones. Data were collected using the User Engagement Scale (UES-SF) and Semi-structured Interviews.

The findings indicate that the mobile Manga and Webtoon reading experience is influenced by readers' prior reading habits and structure. Mixed Readers also obtained higher mean scores than both Digital Manga Readers and Webtoon Readers across most questionnaire dimensions, suggesting that users with experience in both formats are better able to adapt to different reading interfaces and narrative styles. The reading experiment further showed that the narrative structure of Manga had a significant main effect on Perceived Usability (PU) ($F = 3.54$, $p = .036$).

[1] Corresponding author. No. 1, Sec. 3, Zhongxiao E. Rd., Taipei City 10608, Taiwan. zmcdesign@gmail.com

Mixed Composition also showed better overall user experience and higher preference rankings for the Webtoon format. These findings suggest that transforming Manga narrative structures into different presentation formats can influence users' reading experience and preferences, providing design references for future Digital Manga interfaces and panel extraction strategies.

Keywords: *Manga Structure*、*Webtoon*、*Panel Analysis*、 *Manga Digital Transformation*

1 INTRODUCTION

Since the 20th century, Manga has developed as a comic form serialized in print magazines and weekly publications, shaping a highly influential reading culture (Frederik L. Schodt, 1996). Through the arrangement of Panels, page-turning rhythm, and page layout, Manga has been regarded as a mature form of Sequential Art (Will Eisner, 1985). However, with the widespread adoption of smartphones, the ways in which people access information and engage in reading have gradually shifted toward digital media. According to the Digital 2024: Taiwan report, 94.9% of internet users in Taiwan primarily access the internet via smartphones, indicating that mobile devices have become a dominant reading medium in contemporary contexts (*Digital 2024*, 2024). Prior studies further indicate that comics are commonly categorized as a form of digital leisure reading, with reading motivations centered on entertainment and enjoyment, and highly dependent on device convenience and interaction fluency (Loh & Sun, 2022).

Within this context, Webtoon has rapidly emerged as the mainstream format of digital comics due to its alignment with smartphone interaction paradigms (Cho, 2021). By adopting a single reading flow, continuous vertical scrolling, and multimodal elements such as sound, Webtoon reshapes comic narrative rhythm and reading experience (Jang & Song, 2017). In contrast, Manga on smartphones is predominantly presented through scanning or zooming-based formats, which reduce the interactivity inherent to digital forms and, in turn, affect overall reading fluency and immersion (Batinic, 2017). Previous research also suggests that Manga employ diverse narrative structures through different panel arrangements and visual compositions. These structures, however, largely originate from print-based reading logic. Chavanne (2015) regarded Manga as an art form composed of multiple panels and images and proposed three panel composition types: Regular Composition, Semiregular Composition, and Rhetorical Composition. He further introduced Fragmentation as an important principle of panel arrangement to explain how different panel configurations influence the visual relationships between panels, reading paths, and narrative organization (Chavanne, 2015).

Based on this context, this study employed questionnaires and reading tasks to examine readers' preferences for Digital Manga formats and to evaluate whether transforming page-based Manga into Webtoon structures enhances mobile reading experience, thereby informing future Manga digitalization design and applications.

2 METHODS

2.1 Phase One: User Survey

In this phase, the questionnaire targeted participants aged 18 years or older who had used a smartphone to read Digital Manga or Webtoon within the past six months. The questionnaire was adapted from the Digital Reading Situation Scale (Tsai, 2011) and included separate sections for Digital Manga and Webtoon reading formats. In addition to collecting demographic information, reading frequency, and preferred reading format, the questionnaire also examined whether users' reading experiences were influenced by different comic formats (Digital Manga and Webtoon).

2.2 Phase Two: Manga Reading Experiment

In this phase, 30 participants were recruited from the survey respondents to participate in a controlled laboratory experiment. The experiment was conducted using an iPhone 16 Pro with a 6.27-inch (15.93 cm) display as the reading device. Based on an analysis of contemporary serialized Manga, three narrative structures were identified for developing the experimental prototypes: Fragmentation (Semiregular Composition with Fragmentation), characterized by structured square panel layouts arranged in a simple and orderly grid; Rhetorical Composition, characterized by irregular panel layouts; and Mixed Composition, which combines features of both structures. Using Figma, each narrative structure was transformed into two presentation formats, Digital Manga and Webtoon, resulting in six experimental prototypes (A-P, A-W, B-P, B-W, C-P, and C-W). Each prototype consisted of a complete battle-oriented story segment selected from a serialized Manga. After completing the reading task, participants completed the questionnaire, evaluating their reading experience primarily based on the panel layout and composition (Table 1).

Table 1. Six experimental prototypes (with blurred processing applied)

A-P **Fragmentation * Digital Manga**	**B-P** **Rhetorical Composition * Digital Manga**	**C-P** **Mixed Composition * Digital Manga**
		
A-W **Fragmentation * Webtoon**	**B-W** **Rhetorical Composition * Webtoon**	**C-W** **Mixed Composition * Webtoon**
		

Note. A = Frieren: Beyond Journey's End； B = SAKAMOTO DAYS； C = SPY×FAMILY

To ensure the validity of the transformation from Manga to the Webtoon format, this study invited two Webtoon experts with more than one year of professional experience and publishing backgrounds to validate the experimental prototypes. The validation included 30 evaluation items, including panel layout, reading order, text placement, image processing, and overall consistency. Percentage Agreement was adopted as the validity indicator, and the results showed an agreement rate of 86.7%. In addition, interviews were conducted to collect the experts' suggestions regarding the prototypes. The findings indicated that the reading flow and panel arrangement of Webtoons can be adjusted according to the narrative rhythm. Furthermore, both experts emphasized that the gutters between panels play a critical role in the process of transforming Manga into Webtoons, significantly influencing the reading experience. Therefore, the expert feedback was incorporated into the discussion of the subsequent experimental results in this study.

2.2.1 Experimental Procedure

During this phase, the experiment adopted a two-stage design. After reading the prototypes on a smartphone, participants first completed comprehension questions to assess their understanding of the story, followed by the UES-SF questionnaire. The UES-SF consists of 12 items and was used to evaluate differences in reading experience across the experimental prototypes in terms of focused attention, perceived usability, aesthetic appeal, and reward factor (O'Brien et al., 2018). This procedure was repeated three times, with each participant reading three different experimental prototypes presented either as Digital Manga or Webtoon (Figure 1).. After completing all experimental tasks, semi-structured interviews were conducted to explore participants' reading preferences.

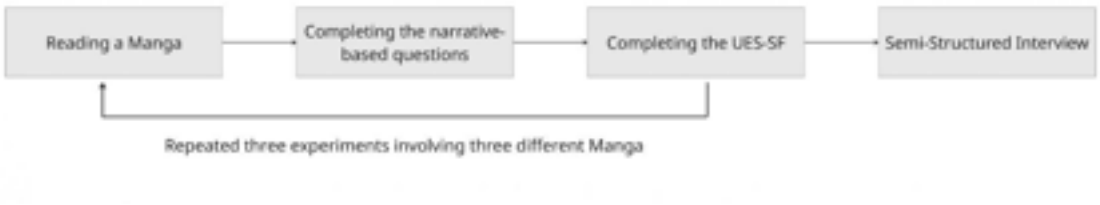


Figure 1. Reading Experimental Procedure

3 RESULT

3.1 Survey Analysis

A total of 149 valid questionnaires were collected. In terms of demographic characteristics, the majority of participants were female (n = 102, 68.5%), while males accounted for 47 participants (31.5%). Regarding age distribution, participants aged 20–29 constituted the largest proportion of the sample (89.3%), followed by those aged 30–39 (8.1%). This demographic distribution indicates that the sample primarily consisted of digital readers in their twenties.

Overall, participants demonstrated substantial reading experience with both Digital Manga and Webtoons. For Digital Manga, 58.4% of participants reported more than five years of reading experience. The most common reading frequency was "once per month" (24.8%), followed by

"once every six months" (22.8%), with typical reading sessions lasting 30–60 minutes (40.9%). For Webtoons, 50.3% of participants reported more than five years of reading experience, and 19.5% indicated that they read Webtoons almost daily. The predominant session length was also 30–60 minutes (32.9%). Overall, participants tended to engage in relatively short reading sessions, with 30–60 minutes being the most common duration for both Digital Manga and Webtoon reading.

3.1.1 Digital Manga & Webtoon Readers Usage Analysis

This study employed a 0–100 scale (with 10-point intervals) to measure participants' relative reading proportions across different comic formats. Lower scores indicated a stronger preference for Digital Manga, whereas higher scores indicated a stronger preference for Webtoons. Based on these scores, participants were classified into three groups: Digital Manga Readers (0–30; n = 35, 23.5%), Webtoon Readers (70–100; n = 56, 37.6%), and Mixed Readers (40–60; n = 58, 38.9%), who frequently read both Digital Manga and Webtoons.

To examine differences in reading experiences among the three reader groups when reading Manga on smartphones, one-way ANOVA was conducted. The results indicated that, when reading Digital Manga on smartphones, significant differences were observed across reading groups in screen size (F = 4.91, p = .009) and ease of interface operation (F = 5.31, p = .006). In addition, for reading fluency (F = 5.10, p = .007), reading comfort (F = 4.73, p = .010), and font size (F = 5.34, p = .006), Mixed Readers reported significantly higher ratings than Webtoon Readers. The results indicated that, compared with Webtoon Readers, Mixed Readers reported higher satisfaction with reading comfort and text presentation when reading Digital Manga on smartphones (Table 2).

Table 2. Differences in Mobile Digital Manga Reading Experience Among Different Reader Groups

Variable	Group 1 (n=35)	Group 2 (n=56)	Group 3 (n=58)	*F*	*p*	Post-hoc (Scheffé)
Planning of content layout	3.46 (0.98)	3.29 (0.89)	3.62 (0.90)	1.92	.151	n.s
Diverse reading experiences	2.91 (0.82)	3.05 (0.86)	3.03 (0.99)	0.28	.755	n.s
Unique reading experience	2.94 (0.84)	2.66 (0.81)	2.98 (0.96)	2.14	.121	n.s
Screen size	3.14 (1.06)	2.63 (0.91)	3.09 (0.82)	4.91**	.009	1,3>2
Reading comfort	3.20 (1.13)	2.80 (0.80)	3.29 (0.80)	4.73*	.01	3 > 2
Font size	2.80(1.08)	2.45(0.87)	3.02(0.91)	5.34**	.006	3 > 2
Text and image resolution	3.54 (0.82)	3.45 (0.76)	3.36 (0.83)	0.56	.571	n.s
Reading fluency	3.69(0.83)	3.20(0.98)	3.71(0.94)	5.10**	.007	3>2
System response speed	3.49 (0.74)	3.20 (0.96)	3.43 (0.99)	1.36	.26	n.s
Ease of interface operation	4.00(0.77)	3.48(0.91)	3.93(0.88)	5.31**	.006	1,3>2

Note. Group 1 = Digital Manga Readers; Group 2 = Webtoon Readers; Group 3 = Mixed Readers.* p < .05. **p < .01.

In the context of reading Webtoons on smartphones, Webtoon Readers reported significantly higher ratings for a unique reading experience (F = 7.02, p = .001) and screen size (F = 5.07, p = .007) than Digital Manga Readers. For reading comfort (F = 10.70, p < .001), font size (F = 6.52, p = .002), Planning of content layout (F = 13.14, p < .001), and Diverse reading experiences (F = 6.41, p = .002), both Webtoon Readers and Mixed Readers reported significantly higher ratings than Digital Manga Readers (Table 3).

Table 3. Differences in Mobile Webtoon Reading Experience Among Different Reader Groups

Variable	Group 1 (n=35)	Group 2 (n=56)	Group 3 (n=58)	*F*	*p*	Post-hoc (Scheffé)
Planning of content layout	3.40(0.88)	4.25(0.61)	3.97(0.84)	13.14***	.000	2,3>1
Diverse reading experiences	3.17(0.99)	3.86(0.92)	3.76(0.90)	6.41**	.002	2,3>1
Unique reading experience	3.34(0.80)	4.02(0.77)	3.72(0.91)	7.02**	.001	2>1
Screen size	3.71 (0.71)	4.18 (0.64)	3.95 (0.71)	5.07**	.007	2>1
Reading comfort	3.57 (1.01)	4.32 (0.58)	4.09 (0.73)	10.70***	.000	2,3>1
Font size	3.74(0.78)	4.27(0.57)	4.14(0.74)	6.52**	.002	2,3>1
Text and image resolution	4.06 (0.73)	4.36 (0.70)	4.22 (0.68)	2.01	.138	n.s
Reading fluency	3.74(0.82)	4.02(0.77)	3.72(0.91)	7.85**	.001	2>1,3
System response speed	3.69 (0.68)	4.02 (0.80)	3.81 (0.87)	2.03	.135	n.s
Ease of interface operation	4.00 (0.73)	4.34 (0.64)	4.17 (0.78)	2.47	.089	n.s

Note. Group 1 = Digital Manga Readers; Group 2 = Webtoon Readers; Group 3 = Mixed Readers. * p < .05. ** p < .01. *** p <0.001.

3.2 Manga Reading Experiment

Based on the mean scores of the UES-SF five-point Likert scale, all six prototypes demonstrated moderately high mean scores across all four dimensions. Among the six prototypes, Mixed Composition × Digital Manga (C-P) achieved the highest mean scores across most dimensions, whereas Fragmentation × Digital Manga (A-P) obtained the lowest mean scores (Table 4).

Table 4. Mean UES-SF Scores for Webtoon and Digital Manga Across Four Dimensions

Dimension	A-P	A-W	B-P	B-W	C-P	C-W
FA	3.69	4.2	4.18	4.24	4.36	4.18
PU	3.09	3.69	3.8	3.93	4.07	3.8
AE	3.84	3.84	4.07	4.22	4	3.93
RW	3.87	4.11	4.09	4	4.24	3.98
Overall Score	3.62	3.96	4.03	4.1	4.17	3.97

Subsequently, Linear Mixed Models (LMMs) were employed to examine the effects of Manga structure, presentation format, and their interaction on the four UES-SF dimensions. For Perceived Usability (PU), a significant main effect of Manga structure was observed (F = 3.54, p = .036), indicating that PU scores differed across Manga structures. No significant main effects of Manga structure were found for Focused Attention (FA), Aesthetic Appeal (AE), or Reward Factor (RW). Furthermore, neither presentation format nor the interaction between Manga structure and presentation format had significant effects on any of the four UES-SF dimensions. These findings indicate that presentation format did not have a significant effect on users' reading experience across the four UES-SF dimensions (Table 5).

Table 5. Linear Mixed Model (LMM) Results of the UES-SF Across Different Formats and Structures

Dimension	Effect	df	*F*	*p*
Focused Attention (FA)	Structures	62.93	1.912	.156
	Formats	72.53	0.963	.330
	Structures * Formats	62.92	2.196	.120
Perceived Usability (PU)	Structures	58.08	3.536*	.036
	Formats	83.39	0.760	.386
	Structures * Formats	58.08	1.962	.150
Aesthetic Appeal (AE)	Structures	63.48	1.563	.218
	Formats	75.07	0.046	.831
	Structures * Formats	63.48	0.294	.746
Reward Factor (RW)	Structures	61.65	0.236	.790
	Formats	80.41	0.061	.805
	Structures * Formats	61.65	0.993	.376

Note. *p<0.05.

Bonferroni-adjusted post hoc pairwise comparisons indicated that Mixed Composition yielded significantly higher Perceived Usability (PU) scores than Fragmentation (mean difference = 0.544, p = .048).No other pairwise comparisons among the Manga structures were statistically significant (Table 6).

Table 6. Post Hoc Pairwise Comparisons of Different Manga Structures on the PU

Comparison I vs J	Mean Difference(I-J)	SE	*p*	Post-hoc (Bonferroni)
A vs B	−0.478	0.224	.112	n.s
A vs C	−0.544	0.219	**.048***	**A<C**
B vs C	−0.067	0.213	1.000	n.s

Note. A = Fragmentation; B = Rhetorical Composition; C = Mixed Composition. *p<0.05.

3.2.1 Analysis of Interface Preferences and Ranking Results Among All Participants

The Friedman test revealed significant differences in preference rankings among the six experimental prototypes ($\chi^2 = 14.99$, $p = .010$), indicating that overall preferences differed across the six prototypes. Among the six experimental prototypes, the Webtoon version of the Mixed Composition (C-W) obtained the lowest mean rank (2.93), indicating the highest overall preference. In contrast, the Digital Manga version of the Fragmentation structure (A-P) obtained the highest mean rank (4.60), indicating the lowest overall preference (Table 7).

Table 7 Preference Rankings of Six Prototypes Structures and Friedman Test Results

Variable	Mean Rank	Rank Order	Statistical Summary
C-W	2.93	1	
B-P	3.13	2	
A-W	3.33	3	N = 30
B-W	3.33	4	$\chi^2 = 14.99$
C-P	3.67	5	df = 5
A-P	4.6	6	$p = .010$*

Note1 . *p<0.05.

3.3 Participant Interviews

The verbatim transcripts from the semi-structured interviews were analyzed using thematic analysis to identify key themes related to the effects of different layout formats on mobile reading experiences. The analysis revealed three primary themes: reading flow, information density, and immersion. Participants consistently emphasized that appropriate gutter spacing in Webtoons helps maintain the reading rhythm, whereas excessive gutter spacing disrupts narrative continuity. Device operability also emerged as a recurring theme, as reading Digital Manga on smartphones often requires frequent zooming, which interrupts the overall reading experience.

4 DISCUSSION

This study investigated the effects of transforming page-based Manga into different digital reading formats for smartphones on users' reading experience. Drawing on Donald Norman's Emotional Design theory, the findings are interpreted through the three levels of visceral, behavioral, and reflective design to explain how different Manga structures and digital reading formats shape users' reading experiences and perceptions (Norman, 2004).

Visceral Level: The mobile reading experience of Digital Manga and Webtoons was jointly influenced by readers' prior reading habits, Manga structure, and digital reading format. The preliminary questionnaire indicated that readers' prior reading experience affected their evaluations of different reading formats. Furthermore, Mixed Readers obtained higher mean scores than both Digital Manga Readers and Webtoon Readers across most questionnaire dimensions. These findings suggest that users with experience reading both Digital Manga and

Webtoons are better able to adapt to different reading interfaces and narrative styles and are less likely to experience a decline in their overall reading experience when the reading format changes. They further indicate that experience with multiple comic reading formats enhances users' adaptability to different digital reading formats.

Behavioral Level: The reading experiment further demonstrated that users' reading experience was primarily influenced by Manga structure and panel composition. Among the three Manga structures, Mixed Composition demonstrated better overall performance in user experience, Perceived Usability (PU), and preference rankings in the Webtoon format, suggesting that appropriately incorporating irregular panel layouts into Manga structure can enhance the mobile reading experience. In contrast, Fragmentation received lower evaluations in both Perceived Usability (PU) and preference rankings in the Digital Manga format. Combined with the interview findings, the results further suggest that panel arrangement and gutter spacing influence users' reading perceptions and overall reading experience.

Rather than simply digitizing printed pages, Manga should be transformed for smartphone reading by restructuring panel arrangements to accommodate scrolling-based interaction. Design strategies such as panel extraction and panel reorganization may provide a more effective approach for adapting Manga to digital reading, offering valuable design implications for the future publication and distribution of Digital Manga.

Reflective Level: Overall, the findings suggest that the effectiveness of digital Manga transformation depends on the underlying Manga structure. Future research could further investigate digital transformation strategies by incorporating additional Manga design elements, such as speech balloons, visual effects, and panel transitions. According to feedback from both interview participants and expert reviewers, excessive gutter spacing in combat scenes may increase scrolling distance when Digital Manga is transformed into the Webtoon format, thereby disrupting reading fluency. This finding supports the argument that panel spacing influences readers' perception of narrative pacing and immersion (Dewey, 2022). The present study extends this perspective by demonstrating that gutter spacing also influences scrolling behavior and reading continuity during smartphone Webtoon reading. Therefore, gutter spacing should be considered a critical design parameter when transforming page-based Manga into Webtoon formats.

These findings further suggest that effective digital transformation should be tailored to the narrative structure and panel composition of Manga, rather than applying a one-size-fits-all conversion approach.

5 CONCLUSION

This study investigated how the transformation of Digital Manga into Webtoon formats for smartphone reading influences users' perception and reading experience through the perspective of Emotional Design Theory. The findings demonstrate that the effectiveness of Digital Manga-to-Webtoon transformation depends on Manga structure, panel composition, and digital reading design. Rather than simply converting printed page layouts into scrolling interfaces,

effective transformation should adapt panel arrangement and gutter spacing according to different Manga structures to better support smartphone reading experiences. These findings provide practical design implications for the future digital transformation of Manga and offer a foundation for future research on mobile comic reading interfaces and Webtoon design.

REFERENCES

Batinic, J. (2017). "Enhanced Webcomics1": An Exploration of the Hybrid Form of Comics on the Digital Medium. *Image & Narrative*, *17*(5). https://www.imageandnarrative.be/index.php/imagenarrative/article/view/1384

Chavanne, R. (2015). The Composition of Comics. *European Comic Art*, *8*(1), 111–144. https://doi.org/10.3167/eca.2015.080108

Cho, H. (2021). The Platformization of Culture: Webtoon Platforms and Media Ecology in Korea and Beyond. *The Journal of Asian Studies*, *80*, 1–21. https://doi.org/10.1017/S0021911820002405

Dewey, M. (2022). Pondering Pauses: Experiencing Stillness in Scrolling Webcomics. *Honors Projects*. https://digitalcommons.spu.edu/honorsprojects/171

Digital 2024: Taiwan. (2024, February 23). DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2024-taiwan

Schodt, F. L. (1996). *Dreamland Japan Writings On Modern Manga*. http://archive.org/details/dreamland-japan

Jang, W., & Song, J. E. (2017). Webtoon as a New Korean Wave in the Process of Glocalization. *Kritika Kultura*, (29). https://doi.org/10.13185/KK2017.02908

Loh, C. E., & Sun, B. (2022). The impact of technology use on adolescents' leisure reading preferences. *Literacy*, *56*(4), 327–339. https://doi.org/10.1111/lit.12282

Norman, D. (2004). Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books.

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, *112*, 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

Tsai, P. (2011). *A Study of Situation and Flow Experience on Digital Reading* [National Taiwan University of Arts]. https://hdl.handle.net/11296/24aj7d

Eisner, W. (1985). *Will Eisner Theory of Comics and Sequential Art*. http://archive.org/details/Will_Eisner_Theory_of_Comics_and_Sequential_Art

INVESTIGATION OF SCREEN PLACEMENT IN MR TO SUPPORT EXOSCOPE SURGERY

Aika Maruyama[*a], Kanta Tuyuki[a] , Kenta Nagai[b], Jue Zhang[a]

[a] *Kansei Information Laboratory, Kogakuin University, Japan*
[b] *Department of Neurosurgery, Tokyo Medical University Ibaraki Medical Center, Japan*

ABSTRACT

In exoscope surgery, monitor placement for presenting the surgical field has been reported to influence the surgeon's gaze movement and physical workload. To address this issue, mixed reality (MR) technology, which enables interaction with virtual information while perceiving the real world, has attracted attention as a flexible information presentation method in medical applications, particularly in neurosurgical support. This study investigated the effects of monitor placement in an MR-based simulated exoscope surgical environment on task performance and visual burden.

A virtual monitor was presented in an MR environment using a head-mounted display, and participants performed an object transfer task using a laparoscopic surgery training kit. Five monitor placement conditions were examined: front, upper, lower, left, and right. Task completion time, eye-tracking data, and subjective evaluations were collected.

The results showed that the lower and front placements provided higher ratings for ease of viewing and concentration, with lower levels of fatigue and discomfort. In contrast, upper and lateral placements increased gaze movement and subjective burden. These findings suggest that flexible virtual monitor placement using MR technology can effectively support exoscope surgery and reduce surgeon workload.

Keywords: *MR, Medical care, Surgical support*

[*] Corresponding author: Kansei Information Laboratory, Master's Program in Informatics, Graduate School of Engineering, Kogakuin University, 1-24-2 Nishishinjuku, Shinjuku-ku, Tokyo 163-8677, Japan. E-mail: jx22325@g.kogakuin.jp

1 INTRODUCTION

High-resolution exoscope systems that combine the advantages of microscopes and endoscopes are increasingly used in neurosurgery. Conventional surgical microscopes require surgeons to maintain a fixed posture while viewing the surgical field through eyepieces, often causing physical strain. Modern exoscope systems display magnified images captured by external cameras on large monitors, allowing flexible positioning and image sharing among multiple personnel. This reduces physical burden and may shorten surgical time [1][2]. However, monitor placement in exoscope systems can alter surgeons' posture and gaze movement, potentially causing discomfort and reduced work efficiency depending on the placement [3][4][5]. Therefore, optimization of monitor placement remains an important issue.

This study employs Mixed Reality (MR) technology to construct a simulated surgical task environment. MR enables interaction with virtual information while perceiving the real world and has been applied in various fields, including medicine [6][7]. In neurosurgery, MR has attracted attention for improving visibility and supporting surgical tasks [8][9].

Using a head-mounted display (HMD), this study presents a monitor in an MR environment and investigates how different monitor placements affect task performance and smoothness during hand–eye coordination tasks.

2 EXPERIMENTAL METHOD

2.1 Experimental Overview

Participants transferred an object from one container to another using a laparoscopic surgery training kit (hereafter, "surgical kit") while observing internal images of the kit displayed on a monitor in the MR space (Figure 1). The external appearance of the surgical kit is shown in Figure 2. Task performance and subjective evaluations were compared across different monitor placement conditions.



Figure 1. Example of the internal image of the surgical kit

Twenty-two participants (13 males and 9 females) took part in the experiment. Seventeen participants were aged 20–24 years, and five participants were aged 25 years or older. Seventeen

participants reported normal or good visual acuity (including corrected vision), whereas five reported reduced visual acuity. Nine participants had prior MR experience, whereas thirteen had no prior MR experience. All participants were right-handed and had no prior experience using a laparoscopic surgery training kit. The study was conducted in accordance with ethical guidelines and approved by an institutional ethics review board.

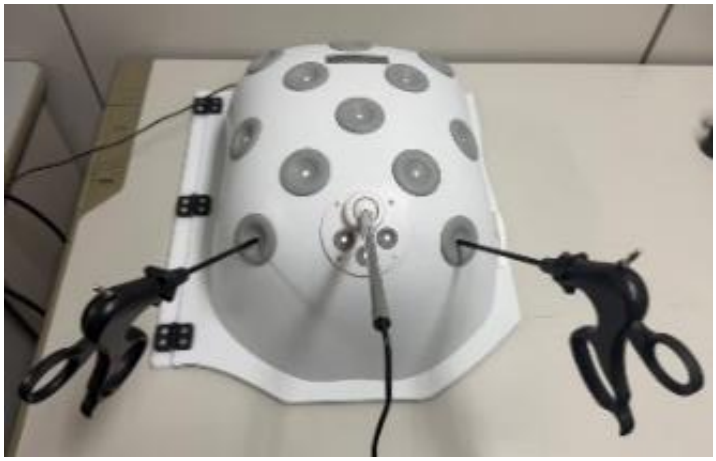


Figure 2. External appearance of the surgical kit

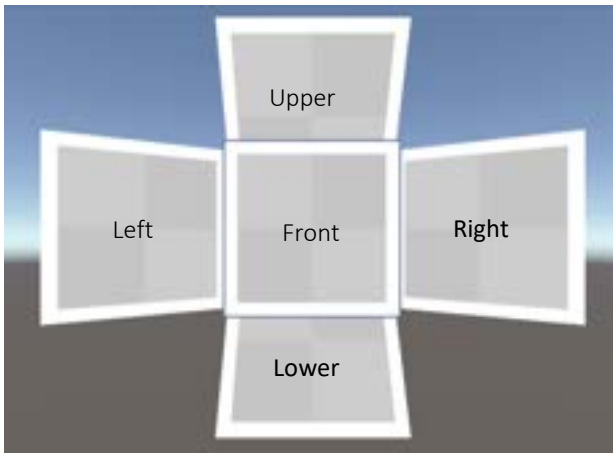


Figure 3. Monitor placement positions

The experimental system was developed in Unity using C# and implemented on a Meta Quest Pro HMD. Real-time internal images of the surgical kit were captured by an external camera, processed using OBS, and streamed via VDO.Ninja, allowing continuous observation during task execution. Five monitor placement conditions were defined: front, upper, lower, right, and left (Figure 3). Based on ergonomic considerations of gaze angle and posture, placements were limited to within ±30° in both vertical and horizontal directions [10]. Eye-tracking data were recorded using the HMD's built-in eye tracker. Gaze data were converted into Unity world coordinates and used to visualize gaze trajectories for qualitative evaluation of gaze behavior under each placement condition.

2.2 Experimental Procedure and Evaluation Items

Participants were informed of the experimental purpose and procedures, provided informed consent, and performed the task while wearing the HMD. To minimize confounding effects of

head movement, participants were instructed to maintain a forward-facing posture, allowing only minimal head motion necessary for task execution. This constraint was intended to isolate the effects of monitor placement on gaze direction and to simulate surgical situations in which gaze and hand positions do not coincide. Under each of the five placement conditions, participants performed the object transfer task while viewing the monitor. Rest periods were provided as needed.

Objective evaluations included task completion time and eye-tracking data. Subjective evaluations were conducted after each condition using a five-point Likert scale assessing ease of viewing (Evaluation 1), ease of concentration (Evaluation 2), fatigue (Evaluation 3), and discomfort (Evaluation 4). For fatigue and discomfort, participants also selected possible causes. After completing all conditions, participants ranked the placements by ease of viewing and task performance, and post-experiment interviews were conducted to collect supplementary comments.

3 RESULTS

One-way ANOVA revealed significant main effects of monitor placement for all subjective evaluation items (Table 1). Multiple comparisons showed that lower placement was rated significantly higher for ease of viewing and concentration than upper and lateral placements, while front placement was rated higher than upper placement. In contrast, upper placement showed significantly higher fatigue and discomfort ratings than lower and front placements (Table 2). No significant differences in task completion time were observed among placement conditions ($p = 0.985$). However, mean task completion times tended to be shorter for front, left, and lower placements and longer for upper and right placements. Gaze trajectory analysis (Figure 4) showed that gaze was concentrated within a narrower area for lower and front placements, whereas upper and lateral placements resulted in wider gaze movement and more frequent gaze shifts.

Table 1. Results of one-way ANOVA for each subjective evaluation item

Evaluation		p-values
Evaluation1	Do you feel that the content on the presented monitor (image) is easy to see?	0.002**
Evaluation2	Does the monitor placement contribute to ease of concentration during the task?	0.000***
Evaluation3	Do you feel fatigue due to the monitor placement?	0.000***
Evaluation4	Do you feel discomfort (unnaturalness or uneasiness) with the monitor placement?	0.000***
	Task completion time	0.985

* $p < .05$, ** $p < .01$, *** $p < .001$

To examine whether the four subjective evaluation items represented independent constructs or reflected common psychological dimensions, Pearson correlation analysis and principal component analysis were conducted using all subjective evaluation data across the five monitor placement conditions. Pearson correlation analysis revealed significant correlations among all four subjective evaluation items (Table 3). Ease of viewing was positively correlated with ease of

concentration (r = 0.724, p < .001) and negatively correlated with fatigue (r = –0.513, p < .001) and discomfort (r = –0.624, p < .001). Fatigue was positively correlated with discomfort (r = 0.580, p < .001). Principal component analysis showed that the four subjective evaluation items reflected a common underlying psychological construct. A single principal component explained 64.8% of the total variance, with high positive loadings for ease of concentration (0.926) and ease of viewing (0.765) and negative loadings for fatigue (–0.667) and discomfort (–0.840) (Table 4).

Table 2. Results of multiple comparisons between monitor placement conditions (p-values)

		Evaluation 1	Evaluation 2	Evaluation 3	Evaluation 4
Front	Upper	0.878	0.006**	0.044*	0.000***
	Lower	0.136	0.793	1.000	1.000
	Right	0.136	0.020*	0.229	0.000***
	Left	0.878	0.020*	0.034*	0.003**
Upper	Lower	0.030*	0.000***	0.003**	0.000***
	Right	0.878	1.000	1.000	1.000
	Left	0.878	1.000	1.000	1.000
Lower	Right	0.000***	0.000***	0.032*	0.000***
	Left	0.009**	0.000***	0.002**	0.000***
Right	Left	0.878	1.000	1.000	1.000

* p < .05, ** p < .01, *** p < .001

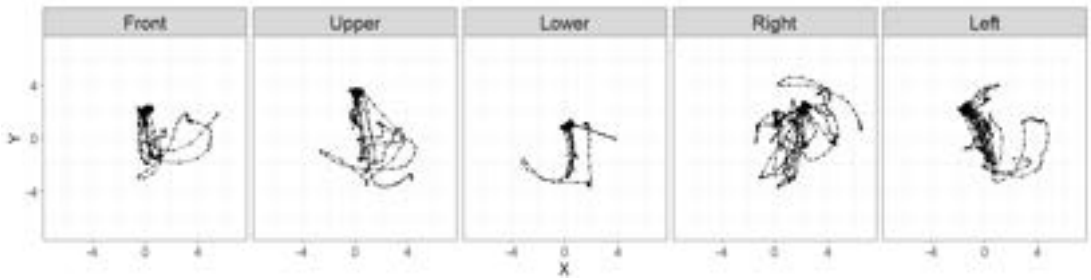


Figure 4. Example of gaze trajectories by monitor placement condition

Table 3. Pearson correlation coefficients among subjective evaluation items

	Viewing	Concentration	Fatigue	Discomfort
Viewing	1.000	0.724***	-0.513***	-0.624***
Concentration		1.000	-0.598***	-0.780***
Fatigue			1.000	0.580***
Discomfort				1.000

*** p < .001.

Additional analyses were conducted according to participant characteristics. Because sex and prior MR experience consisted of two groups, these variables were compared using independent-samples t-tests. In contrast, age and visual acuity were examined descriptively because of the relatively small and unequal subgroup sizes. Independent-samples t-tests revealed no significant

differences according to prior MR experience for any evaluation item or monitor placement condition. Likewise, no significant sex differences were found except for fatigue under the lower placement condition ($p = 0.046$). Descriptive comparisons indicated similar trends across the age and visual acuity groups.

Table 4. Principal component loadings of subjective evaluation items

Item	Factor1
Concentration	0.926
Discomfort	-0.840
Viewing	0.765
Fatigue	-0.667

Eigenvalue = 2.594. Variance explained = 64.8%.

4 DISCUSSION

The results indicate that monitor placement in an MR environment significantly influences visual comfort, concentration, fatigue, and discomfort during task execution. Lower placement provided the highest usability with minimal fatigue and discomfort, likely due to reduced gaze movement and consistency with natural downward viewing during manual tasks. Front placement was also effective but showed slightly increased vertical gaze movement and fatigue. Upper and lateral placements increased gaze travel distance, leading to greater neck and ocular strain, as reflected in higher fatigue and discomfort ratings. The lack of significant differences in task completion time may be attributed to task simplicity and learning effects across trials. Correlation and principal component analyses further indicated that the four subjective evaluation items reflected a common underlying psychological construct rather than independent dimensions. This finding suggests that improvements in monitor placement simultaneously enhance ease of viewing and concentration while reducing fatigue and discomfort, supporting an integrated interpretation of subjective evaluations from a Kansei engineering perspective. Additional analyses according to participant characteristics showed that the effects of monitor placement were generally consistent regardless of sex, age, visual acuity, or prior MR experience. In particular, no significant differences were found between participants with and without prior MR experience for any evaluation item or monitor placement condition. Likewise, no significant sex differences were observed except for fatigue under the lower placement condition. These findings suggest that the observed effects of monitor placement were relatively robust across participants with different backgrounds. However, because the subgroup sizes were limited and not fully balanced, further studies with larger and more diverse participant populations are needed to confirm the generalizability of these findings.

5 CONCLUSION AND FUTURE WORK

This study investigated the effects of monitor placement conditions in an MR environment on work efficiency and task-related burden. The results revealed that the lower and front placements offer superior ease of viewing and concentration with lower levels of fatigue and discomfort. In

contrast, the upper and lateral placements tended to increase gaze movement and subjective burden. Therefore, in support of exoscope surgery, it is desirable to align the surgeon's operation direction with gaze direction as much as possible and to minimize gaze movement through appropriate monitor placement. However, in real surgical environments, it is often difficult to maintain ideal monitor placement due to constraints such as equipment, operating tables, and staff arrangement. In this regard, MR technology enables the construction of virtual environments in which monitors can be flexibly positioned according to the surgeon's stance and operation direction. Consequently, future applications of MR-based surgical support and workload reduction are highly anticipated.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to all participants and to everyone who supported this research in various ways through valuable advice, technical assistance, and helpful discussions.

REFERENCES

[1] Murakami, T., Toyota, S., Suematsu, T., Wada, Y., Nakagawa, K., Shimizu, T., Mori, K., & Taki, T. (2021). Four hands surgery for intracerebral hemorrhage using ORBEYE: Educational values and ergonomic advantages – A technical note. Asian Journal of Neurosurgery, 16(3), 634–637. Retrieved from https://doi.org/10.4103/ajns.AJNS_148_21

[2] Montemurro, N., Scerrati, A., Ricciardi, L., & Trevisi, G. (2022). The exoscope in neurosurgery: An overview of the current literature of intraoperative use in brain and spine surgery. Journal of Clinical Medicine, 11(1), 223. Retrieved from https://doi.org/10.3390/jcm11010223

[3] Takahashi, S., Toda, M., Nishimoto, M., Ishihara, E., Miwa, T., Akiyama, T., Horiguchi, T., Sasaki, H., Yoshida, K., et al. (2018). Pros and cons of using ORBEYE™ for microneurosurgery. Clinical Neurology and Neurosurgery, 174, 57–62. https://doi.org/10.1016/j.clineuro.2018.09.010

[4] Roethe, A. L., Landgraf, P., Schröder, T., Misch, M., Vajkoczy, P., & Picht, T. (2020). Monitor-based exoscopic 3D4k neurosurgical interventions: A two-phase prospective-randomized clinical evaluation of a novel hybrid device. Acta Neurochirurgica, 162(12), 2949–2961. Retrieved from https://doi.org/10.1007/s00701-020-04361-2

[5] Maeda, T., Oigawa, H., Onodera, K., Sato, H., Suzuki, K., & Kurita, H. (2023). Pros and Cons of Exoscopic Intracranial Aneurysm Repair. Surgery for Cerebral Stroke, 51(5), 397–404. Retrieved from https://doi.org/10.2335/scs.51.397.

[6] Shiga, Y., Sugimoto, M., Abe, M., Nishikimi, R., Hoshina, Y., Yoneoka, Y., Saikawa, S., Inoue, Y., Yoshimatsu, T., Kusakabe, M., & Kameyama, S. (2018). A study on urologation surgery applying

mixed reality MR, augmented reality AR, virtual reality VR. Japanese Journal of Endourology, 31(2), 253–259. Retrieved from https://doi.org/10.11302/jsejje.31.253

[7] Magalhães, R., Oliveira, A., Terroso, D., Vilaça, A., Veloso, R., Marques, A., Pereira, J., & Coelho, L. (2024). Mixed reality in the operating room: A systematic review. Journal of Medical Systems, 48(1), 76. Retrieved from https://doi.org/10.1007/s10916-024-02095-7

[8] Noguchi, H. (2021). Current status of VR, AR, and MR in medical applications. Iryo Kiki Gaku (Journal of Medical Instrumentation), 91(5), 462–465. Retrieved from https://doi.org/10.4286/jjmi.91.462

[9] Yoshikane, T., Nagai, H., Hayasi, K. (2024). Mixed reality and head-mounted devices for surgical learning and training. Japanese Journal of Neurosurgery, 33(10), 714–717. Retrieved from https://doi.org/10.7887/jcns.33.714

[10] Japanese Standards Association. (2002). JIS Z 8515: Ergonomics—Office work with visual display terminals (VDTs)—Requirements for workstation layout and working posture. Retrieved from https://kikakurui.com/z8/Z8515-2002-01.html

KANSEI EVALUATION OF JAPANESE MUSIC CONSIDERING ARTIST'S GENERATION AND LISTENER'S AGE

Hiroshi Takenouchi[1a], Masaki Kawakubo[a]

[a] *Fukuoka Institute of Technology, Japan, h-takenouchi@fit.ac.jp*

ABSTRACT

We analyze trends in music listener preferences using Kansei evaluation and taking into account differences in listener age and the generation of the artist. This study conducts Kansei evaluation experiments with participants ranging from their late teens to their forties, who listen to music performed by male dance performance groups considering of experienced and younger artists. We employ impression evaluations for various types of music using the semantic differential method, and we analyze the results through factor analysis and clustering. The results reveal that the factors that express impressions of music by experienced and younger groups differed from each other. In addition, we compared impressions between music-only presentations and music videos that added visual information to the audio, confirming that the impressions did not differ between the two conditions.

Keywords: *Kansei evaluation, Semantic differential method, Japanese pops, Multivariate analysis.*

1 INTRODUCTION

Musical trends (genres and styles) change over time. Such changes stem from differences in the ways that listeners of different age groups perceive music, as well as variations in expression influenced by generational differences among the artists who sing or perform it. Therefore, even when artists belong to the same category—such as dancer performers or idols—the impressions of their music can differ depending on the artist' and the listeners' generation.

Previous studies of the impression analysis of music have included the Kansei evaluation of key modulation in classical music (Hasegawa, 2025) and in background music (BGM) (Kampfer, 2010), the analysis of music appreciation in terms of affective evaluation and neural activity (Kakegawa, 2012), and Kansei evaluations of drum rhythms (Kurosaki, 2011). These studies focused on listeners' impressions and demonstrated how music is received: however, they did not analyze musical impressions in relation to generational differences among artists and listeners.

[1]Corresponding author

Considering these differences allow affective evaluations to be analyzed from multiple perspectives—such as how younger listeners perceive the music of experienced artists, or how middle-aged listeners respond to songs by younger performers.

This study conducts Kansei evaluations that consider listener age and artist generation to analyze trends in listeners preferences. We focus on listeners aged from their late teens to their forties. Careful consideration is required when selecting artist generations. For example, when evaluating songs by a younger jazz performer and an experienced rock musician, the fundamental differences in musical atmosphere and genre make direct comparison difficult. Therefore, it is necessary to select both experienced and younger artists whose music belongs to the same genre or category. In this study, we focus on male dance performance groups affiliated with love, dream, happiness (LDH) Japan Inc., which represents a wide range of artists, from emerging talents to seasoned veterans.

In this experiment, Kansei evaluation is conducted using the semantic differential (SD) method (Osgood, 1964), based on debut, most recent, and representative songs of six artist groups (three experienced and three younger groups) by listener participants from their late teens to their forties. The experiment analyzes trends in music appreciation between experienced and younger groups based on the Kansei evaluation results.

Through the experiment, we aim to clarify the following three points:

1. Whether the impression of a song changes depending on the listener's age.
2. Whether the impression of a song changes depending on the artist's generation.
3. Whether the impression of a song changes when visual elements are added to the audio.

Items 1 and 2 are the main focus of this study, involving a comparative investigation of impression evaluation tendencies based on differences in listener age and artist generation. Regarding item 3, since music videos (MVs), which are often released alongside songs, may significantly impact on listener impressions, this study examines differences in impressions when evaluating the audio-only versions and those from MVs that include visual information.

2 EVALUATION EXPERIMENTS

2.1 Outline

We conducted two Kansei evaluations: one to assess listeners' impressions when hearing the music alone, and another to evaluate their impressions when visual information was added in a MV. In Experiment 1, we presented only a 15-second excerpt from the chorus section of each artist's debut and latest songs and examined age-based differences in listener impressions through factor analysis and cluster analysis using the farthest neighbor method. In Experiment 2, we presented each artist's representative song in two formats—audio-only and audio with visual content (MV)—and investigate age-based differences in impressions using factor analysis.

The factor analysis parameters are as follows: the factor extraction method uses bartlett method, the rotation method of factor axis uses Varimax rotation, and the number of factors is

three in the Experiment 1 and two in the Experiment 2 decided by scree plot results of eigen values of the correlation matrix (Kaiser standard). Furthermore, as mentioned in Sections 2.2 and 2.3, the results of this study were analyzed using exploratory factor analysis because the number of participants was relatively small.

Table 1 lists the LDH Japan-affiliated artists and songs selected for both experiments. Shaded cells indicate experienced artists, while unshaded cells represent younger artists. The latest and representative songs shown in Table 1 were selected on October 30, 2024. For the representative songs, those of the experienced groups were independently selected by the author, based on each artist's popular tracks, while those of the younger groups were chosen as the songs with the highest number of views on each artist's official YouTube channel[2]. The artists are labeled A through F, the songs for each are labeled 1 through 3, and in Experiment 2, audio-only and MV conditions are distinguished by designating them as a or b, respectively. For example, the latest song (2) by GENERATIONS (C) is labeled as C2, and the MV (b) for "Lollipop" (3) by LIL LEAGUE (D) is labeled as D3b. Using these songs allows differences in the musical styles of artists from the same agency to be represented according to their debut periods and the selected songs. However, this approach does not control for differences arising from the time of composition or the year of release.

Table 1 Artists music in LDH Japan

Artist / Debut year	Experiment 1		Experiment 2
	Debut song (1)	Latest song (2)	Representative song (3)
EXILE (A) / 2001	Your eyes only	Reason	Rising sun (a, b)
J Soul Brothers III (B) / 2010	Best Friend's Girl	Baby don't cry	R. Y. U. S. E. I (a, b)
GENERATIONS (C) / 2012	BRAVE IT OUT	Endless·Journey	AGEHA (a, b)
LIL LEAGUE (D) / 2022	Hunter	Youth Spark	Lollipop (a, b)
WOLF HOWL HARMONY (E) / 2023	Sweet Rain	ROLLIN' STONES	Frozen Butterfly (a, b)
THE JET BOY BANGERZ (F) / 2023	Jettin'	UNBRAKABLE	PHOTOGENIC (a, b)

Table 2 presents the adjective pairs used in the SD method for both experiments. A total of 10 adjective pairs were chosen to investigate impressions related to the songs and MVs. Responses were rated on a 7-point scale, and the survey was conducted anonymously using Google Forms.

This study was conducted in accordance with the ethical guidelines for research involving human subjects at the Fukuoka Institute of Technology and received official approval (Approval

[2] https://www.youtube.com/@lilleague, https://www.youtube.com/@wolfhowlharmony,

https://www.youtube.com/@THE_JET_BOY_BANGERZ

No. hm18–24). Before participation, all participants were informed of the experimental procedures to be performed and precautions taken, and they provided voluntary informed consent.

Table 2 Adjective pairs for SD method

1 – 7	1 – 7	1 – 7
Powerful – Not powerful	Flashy – Modest	Fun – Sad
Passionate – Calm	Energetic – Quiet	Bright – Dark
Strong – Weak	Cool – Cute	Calm – Intense
		Like – Dislike

2.2 Experiment 1

The participants in the first experiment were divided into three age groups: 10 teenagers, 10 individuals in their twenties, 4 in their thirties, and 6 in their forties, totaling 30 participants (17 males and 13 females). Each participant was asked to evaluate their impressions of both the debut and latest songs of the artists listed in Table 1, using the 7-stage adjective scales shown in Table 2. We analyzed these results to determine how differences in artist generation and listener age influenced the impression ratings of the music.

2.3 Experiment 2

The participants were 22 individuals in their twenties (11 males and 11 females). Each participant was asked to evaluate their impressions of both the audio recordings and the MVs of the representative songs by the artists listed in Table 1, using the 7-point adjective scales shown in Table 2. We analyzed these results to demonstrate how the addition of visual information influenced the impression ratings compared to audio-only presentations.

3 EXPERIMENTAL RESULTS AND DISCUSSION

3.1 Results of Experiment 1

Table 3 presents the correlation coefficients calculated from the mean ratings of each adjective pair. In Table 3, correlation coefficients greater than or equal to 0.8 are highlighted in red, whereas those less than or equal to −0.8 are highlighted in blue. The results indicate that strongly correlated adjective pairs were observed for all listener age groups. This suggests that some evaluation adjectives may be redundant, and therefore further refinement of the evaluation adjectives should be considered in future studies. Table 4 presents the communalities obtained from the factor analysis. In most cases, the communalities exceeded 0.7, indicating that the adjective pairs were well explained by the extracted common factors.

Table 5 presents the factor loadings that were extracted from the analysis. For all listener age groups, the number of factors was set to three based on the changes in eigenvalues that were observed using the scree plot method. The adjective pair “like–dislike” was excluded from the factor analysis, as it reflected evaluative preference. According to Table 5, the cumulative contribution rates for the three factors were 0.88, 0.94, and 0.91 for each age group. The factor loadings highlighted in red indicate the highest loading among the three factors for each adjective

pair. From these results, the three factors in the experiment were defined as shown in the second row of Table 5.

Table 3 Results of correlation matrix (Exp.1)

*Teens Twenties Thirty/Forties	Powerful–Not powerful	Passionate–Calm	Strong–Weak	Flashy–Modest	Energetic–Quiet	Cool–Cute	Fun–Sad	Bright–Dark	Calm–Intense
Powerful–Not powerful	1.00								
Passionate–Calm	0.82	1.00							
	0.80								
	0.76								
Strong–Weak	0.73	0.66	1.00						
	0.96	0.87							
	0.91	0.82							
Flashy–Modest	0.64	0.61	0.68	1.00					
	0.63	0.75	0.70						
	0.80	0.70	0.90						
Energetic–Quiet	0.77	0.60	0.71	0.92	1.00				
	0.79	0.78	0.85	0.87					
	0.82	0.69	0.87	0.92					
Cool–Cute	0.02	0.05	-0.06	-0.50	-0.50	1.00			
	0.18	-0.07	0.08	-0.37	-0.41				
	-0.47	-0.32	-0.49	-0.62	-0.67				
Fun–Sad	0.79	0.63	0.73	0.92	0.97	-0.42	1.00		
	0.76	0.78	0.83	0.79	0.98	-0.43			
	0.76	0.64	0.86	0.98	0.94	-0.70			
Bright–Dark	0.66	0.63	0.60	0.93	0.88	-0.37	0.93	1.00	
	0.67	0.79	0.79	0.79	0.94	-0.51	0.97		
	0.78	0.65	0.87	0.94	0.96	-0.74	0.96		
Calm–Intense	-0.89	-0.79	-0.73	-0.86	-0.93	0.21	-0.93	-0.87	1.00
	-0.93	-0.83	-0.94	-0.76	-0.94	0.11	-0.93	-0.84	
	-0.87	-0.52	-0.84	-0.74	-0.69	0.33	-0.70	-0.74	

Table 4 Results of communalities (Exp.1)

	Powerful–Not powerful	Passionate–Calm	Strong–Weak	Flashy–Modest	Energetic–Quiet	Cool–Cute	Fun–Sad	Bright–Dark	Calm–Intense
Teens	0.92	0.76	0.61	0.92	1.00	0.81	0.97	1.00	0.98
Twenties	0.99	0.75	0.95	1.00	0.99	0.85	1.00	0.95	0.98
Thirties/Forties	0.91	0.95	0.96	0.97	0.91	0.55	1.00	0.93	1.00

The results presented in Table 5 indicate that the responses from listeners in their teens and twenties showed factors that had largely similar tendencies. The contribution rates for each factor and the adjective pairs with high loadings were also comparable across these age groups. In contrast, the responses from listeners in their thirties and forties revealed distinct factors—such as *happiness*, *stability*, and *emotional depth*—that were not observed in the responses from the younger age groups. This may be because most participants in their teens and twenties were university students, whereas those in their thirties and forties were working adults. Since the songs that are commonly listened to during adolescence—a period when people are generally considered to listen to music most frequently—are likely to differ between current university students and working adults, their tendencies in Kansei evaluation of music may also differ.

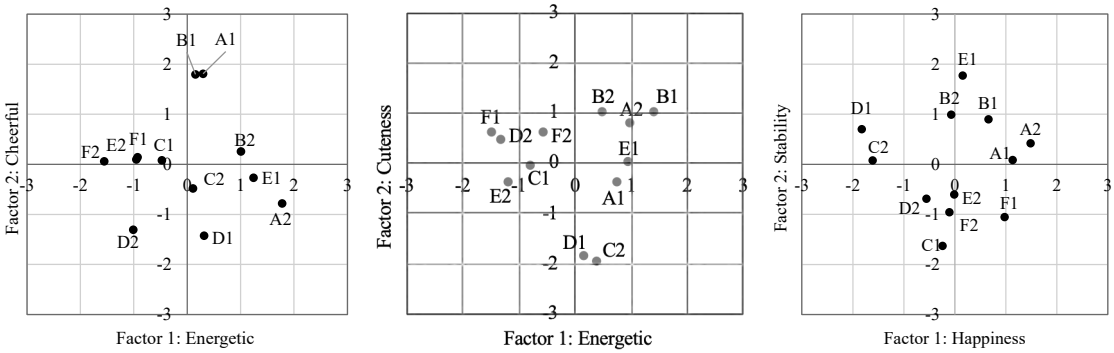
Table 5 Extracted factor loadings

Age group	Teens			Twenties			Thirties/Forties		
Extracted factors	F1: Energetic	F2: Cheerful	F3: Cuteness	F1: Energetic	F2 Cuteness	F3 Cheerful	F1: Happiness	F2: Stability	F3: Emotional depth
Contribution rate	0.46	0.26	0.17	0.61	0.20	0.14	0.40	0.29	0.23
Powerful–Not powerful	0.92	0.27		0.96	–0.17	0.23	0.36	0.70	0.53
Passionate–Calm	0.74	0.40	–0.22	0.77		0.39	0.28	0.27	0.89
Strong–Weak	0.72	0.26	0.13	0.93		0.27	0.50	0.61	0.58
Flashy–Modest	0.53	0.67	0.44	0.54	0.34	0.77	0.78	0.45	0.40
Energetic–Quiet	0.72	0.48	0.51	0.82	0.43	0.37	0.74	0.42	0.43
Cool–Cute		–0.19	–0.88		–0.91	–0.12	–0.73	–0.10	–0.11
Fun–Sad	0.68	0.60	0.39	0.84	0.50	0.23	0.85	0.40	0.33
Bright–Dark	0.49	0.83	0.27	0.74	0.57	0.27	0.77	0.47	0.34
Calm–Intense	–0.84	–0.49	–0.19	–0.94	–0.15	–0.27	–0.30	–0.93	–0.21

Figure 1 presents the mapping of factor scores for each song along with the first and second factors. Across all listener age groups, songs by experienced artists (A1, A2, B1, and B2) were positioned in proximity. In contrast, songs by younger artists (D1, D2, E1, E2, F1, and F2) tended to be located on the opposite side of the plot relative to the experienced songs. Table 6 gives the

mean response values for each song on the "like–dislike" adjective pair. In Table 6, the three lowest mean values are highlighted in red. These results indicate the latest songs tended to be rated as preferable than the debut songs.

Table 7 presents the clustering results based on the mean responses from the SD method for each song. Among listeners in their teens and twenties, the clustering patterns were largely consistent; however, participants in their thirties and forties classified D2 song into a different cluster. Cluster 1 primarily consisted of songs by experienced artists, and Clusters 2 and 3 were predominantly composed of songs by younger artists. The distinct clustering provided for song D2 suggests that artist D could evoke a different impression than artists E and F, despite also being categorized as a younger artist.



(a) Teens (b) Twenties (c) Thirties/Forties

Figure 1 Results of factor scores (Exp.1)

Table 6 Average values for "like – dislike"

	Teens	Twenties	Thirties/ Forties		Teens	Twenties	Thirties/ Forties
A1	3.00	2.11	3.33	A2	3.50	2.25	3.25
B1	3.63	2.00	3.13	B2	3.88	2.88	3.50
C1	3.38	2.13	3.00	C2	2.38	2.38	2.38
D1	1.88	2.25	2.88	D2	3.00	1.88	2.38
E1	3.50	2.25	2.63	E2	3.50	1.63	2.75
F1	3.25	2.25	2.75	F2	2.50	2.25	3.00

Table 7 Results of clustering (clusters: 3)

	Teens	Twenties	Thirties/Forties
Cluster 1	A1, B1, E1, A2, B2	A1, B1, E1, A2, B2	A1, B1, E1, A2, B2
Cluster 2	C1, F1, D2, E2, F2	C1, F1, D2, E2, F2	C1, F1, E2, F2
Cluster 3	D1, C2	D1, C2	D1, C2, D2

3.2 Results of Experiment 2

Table 8 presents the correlation matrix for the adjective pairs. Similar to Table 3, correlation coefficients greater than or equal to 0.8 (less than or equal to −0.8) are highlighted in red (blue). The results indicate a strong tendency for similar Kansei evaluations to be obtained across all songs and MVs. This suggests that the evaluations were similar because the evaluation objects consisted of only two types of stimuli, audio-only and visual (MV) information for each song and that no clear differences in evaluation tendencies attributable to the evaluation objects were observed among the participants. Table 9 presents the communalities obtained from the factor analysis. Similar to the Experiment 1, the communalities exceeded 0.7 in most cases, indicating that the adjective pairs were well explained by the extracted common factors.

Figure 2 presents the mapping of factor scores for each song on the first and second factors. The number of factors in the analysis was set to 2, based on the changes in eigenvalues that were observed using the scree plot method. The adjective pair “like–dislike” was excluded from the analysis because it represented evaluative preference. The contribution rates of the two factors were 0.75 and 0.19, respectively, and they had a cumulative contribution rate of 0.94. Because the factor loadings for each adjective pair closely matched the loadings for the responses of listeners in their twenties in Table 5, the factors were labeled identically: Factor 1 was labeled “Energetic” and Factor 2 “Cuteness.” From the data shown in Figure 2, it was clear that the impressions of the audio-only versions and the MVs of the same songs were largely similar.

Table 8 Results of correlation matrix (Exp.2)

	Powerful – Not powerful	Passionate – Calm	Strong – Weak	Flashy – Modest	Energetic – Quiet	Cool – Cute	Fun – Sad	Bright – Dark	Calm – Intense
Powerful – Not powerful	1.00								
Passionate – Calm	0.95	1.00							
Strong – Weak	0.95	0.99	1.00						
Flashy – Modest	0.94	0.90	0.91	1.00					
Energetic – Quiet	0.95	0.96	0.97	0.92	1.00				
Cool – Cute	0.43	0.58	0.61	0.41	0.55	1.00			
Fun – Sad	0.95	0.92	0.92	0.94	0.92	0.30	1.00		
Bright – Dark	0.93	0.96	0.96	0.93	0.96	0.52	0.96	1.00	
Calm – Intense	-0.91	-0.86	-0.89	-0.87	-0.94	-0.40	-0.86	-0.85	1.00

Table 9 Results of communalities (Exp.2)

Powerful– Not powerful	Passionate– Calm	Strong– Weak	Flashy– Modest	Energetic– Quiet	Cool– Cute	Fun– Sad	Bright– Dark	Calm– Intense
0.95	0.98	1.00	0.90	0.95	0.87	1.00	0.96	0.81

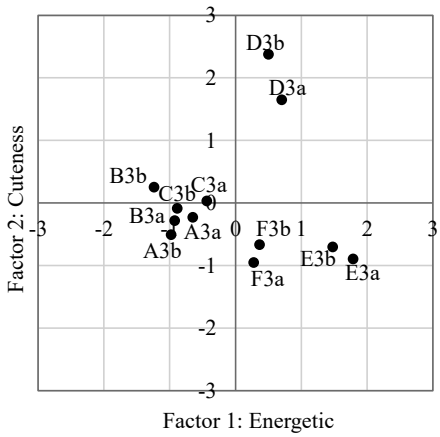


Figure 2 Results of factor scores (Exp.2)

3.3 Discussion

The results of Experiment 1 indicated that the factors extracted from the responses of listeners in their teens and twenties were noticeably different from those of listeners in their thirties and forties. The impressions of the songs may have varied across listener age groups. Factor analysis revealed that songs performed by experienced artists tended to be positioned more closely together, which indicates that listeners across age groups may have had similar impressions of these songs. Furthermore, the clustering results based on the mean responses from the SD method showed consistent classification patterns across all listener age groups. Therefore, the impressions of songs by experienced and younger artists were distinct.

The results of Experiment 2 confirmed that the audio-only and MV versions of the same songs elicited similar impression patterns. While it was hypothesized that the addition of visual information through MVs could alter participants' impressions, this trend was not observed in the experiment.

According to these findings, the results of Experiment 1 support verification item 1 as described in Section 1: there were clear differences in the factors that were extracted from the SD method responses between listeners in their teens and twenties and those in their thirties and forties. Experiment 1 also indicated that songs by experienced and younger artists produced distinct impression patterns, reflecting generational differences in the artists themselves. Finally, verification item 3 was addressed using the results of Experiment 2, which showed that the addition of visual information through MVs did not alter the participants' impressions of the songs. However, it cannot be excluded that the experimental results were affected not only by the limited participant sample but also by factors other than the artists' debut years, such as the musical characteristics of the songs.

3.4 Limitation & Future Research

It cannot be denied that the sample size in this study was relatively small; therefore, the results of the factor analysis should be regarded as exploratory. To examine the robustness of the findings, we conducted a leave-one-participant-out analysis by excluding each of the 10 participants in turn and performing factor analysis on the responses of the remaining nine

participants. The results showed that, for a particular exclusion case, a different factor structure was extracted compared with the other exclusion patterns. Therefore, future studies should include a larger number of participants to improve the robustness and generalizability of the findings. In addition, because the participants in their 30s and 40s were combined solely to balance the group sizes, it was not possible to examine age-related differences in Kansei evaluation among participants aged thirties and older.

Future research will focus on the relationship between musical features and impression ratings. This study focused solely on impression evaluations of songs and MVs, without taking into account the musical features themselves. If specific musical features influence user impressions, this will enable a more detailed analysis of interactions related to users' impression ratings. In addition, we will further investigate the differences in impression ratings between audio-only versions and MVs, including non-musical elements such as filming scenes related to the song's theme and the artist's clothing.

4 CONCLUSION

We conducted Kansei evaluations, taking into account the age of music listeners and generational differences among artists to analyze listeners' preference tendencies. The results showed that impressions of songs varied according to the listener's age group, and that songs by artists of different generations elicited distinct impressions. Moreover, it was suggested that the addition of visual information through MVs did not have an impact on impression ratings. In future, we will map musical attributes to Kansei evaluations to help clarify listener preferences.

REFERENCES

Hasegawa, P. R., Asakura, T. (2025). Reversal of Emotional Impressions through Key Change in Classical Music: An Assessment Using Major and Minor Transpositions. *International Journal of Affective Engineering*, 24(3), 267–277. https://doi.org/10.5057/ijae.IJAE-D-24-00052

Kakegawa, M., Komiyama, R., Masakura, Y., Kikuchi, M. (2012). Analysis of music appreciation by Kansei evaluation and brain activity. In *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligence Systems*, (pp.183–186). https://doi.org/10.1109/SCIS-ISIS.2012.6505407

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2010). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, 39(4), 424–448. https://doi.org/10.1177/0305735610376261

Kurotaki, Y., Shiizuka, H. (2011). Measurement and Classification of the Affective Value of Drum Rhythm Patterns. In *Proceedings of 2011 Third International Conference on Intelligent Networking and Collaborative Systems*, (pp.872–875). https://doi.org/10.1109/INCoS.2011.116

Osgood, C. E. (1964). Semantic Differential Technique in the Comparative Study of Cultures. *American Anthropologist*, 66(3), 171–200. http://www.jstor.org/stable/669329

NOVELTY EVALUATION OF AN IDEATION METHOD UTILIZING REPETITIVE TASKS VIA NOUN METAPHORS

RESEARCH ON IDEATION METHODS FOR DESIGN BEGINNERS (1)

Yoshinori MURAYAMA*[a], Hiroki SATO [b]

*[a]*Chiba Institute of Technology Graduate School, JAPAN*

[b]*Chiba Institute of Technology, JAPAN*

ABSTRACT

Design novices often experience difficulty expressing creativity because of limited access to information. To address this issue, we proposed the Leap Association Method (LAM), an ideation method for generating novel design proposals. LAM requires users to repeatedly generate metaphorical expressions from a given theme, creating difficult-to-interpret noun metaphors that increase the conceptual distance from the original theme. The final associated word is then used as a cue for generating a design proposal.This study examined the process and effectiveness of LAM through two experiments. The results provided partial support for its effectiveness. In the first experiment, 18 design students generated design proposals using LAM. After being presented with a theme, the participants created idea sketches, which were evaluated for novelty by eight design instructors. More than half of the design proposals received high novelty ratings. Interviews were then conducted, and the conceptual distance between the theme and the associated metaphorical expressions was scored. Design proposals with higher novelty ratings showed higher total conceptual-distance scores, and the conceptual-distance score increased with each association.In the second experiment, an improved LAM was compared with the checklist method. Twenty design students generated design proposals using both ideation methods. The results showed no significant difference in the clarity of the ideation method between LAM and the checklist method. However, more than half of the participants evaluated the design proposals generated by LAM as being more novel.These results indicate that LAM can support design novices in generating novel design proposals by progressively increasing conceptual distance through repeated noun-metaphor associations.

Keywords: *Design education, Ideation Methods, Creative Thinking, Metaphor*

**2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan. s24P9508LZ@s.chibatech.ac.jp*

1 INTRODUCTION

The design process rarely proceeds in a fixed sequence. Instead, it often progresses randomly and iteratively. It has been reported that increasing the number of iterations improves the value of design proposals (Kumar, 2015). Such trial-and-error processes have been adopted in many ideation methods. At the same time, the interpretation of design has expanded through approaches such as design thinking, and the number of people studying design has increased. However, constraints on acquiring and using information, such as filter bubbles and echo chambers, are expected to affect creativity, making idea generation during divergent thinking more difficult. For example, free association may limit thinking, while forced association may limit novelty. To address this issue, we previously proposed the Leap Association Method (LAM), an ideation method that enables design novices to break away from fixed conceptual frameworks and generate novel ideas (Murayama & Sato, 2025). LAM is intended for use in design education. By repeatedly generating noun metaphors from a given theme, the method gradually increases the conceptual distance from the original theme. As a result, concepts unrelated to the theme are generated, supporting idea generation with unexpectedness and novelty. This study provides the basis for establishing LAM as an ideation method. It examines the ideation process and the effectiveness of LAM. Future research will further evaluate the method and establish an ideation method for design novices.

2 PURPOSE

The purpose of this study is to evaluate, through two experiments, whether LAM enables design novices to break away from existing conceptual frameworks and produce novel design proposals.

3 PREVIOUS APPROACHES TO IDEATION

According to Watanabe et al. (1992), the conceptual thinking process differs in the early stage of the design process. Because design novices work under limited information, they tend to generate ideas from familiar knowledge. Therefore, it is difficult for them to create unrelated concepts through free association, combine unrelated concepts, or generate ideas through conceptual combination. Tanaka and Noguchi (2004) proposed an ideation method using negative expressions during divergent thinking to break away from existing conceptual frameworks. In this method, multiple negative expressions describing low-level functions of an object are generated and combined through forced inference to produce novel ideas and unexpected images. However, because design novices have limited prior knowledge, they may find it difficult to accept negative interpretations of existing concepts, resulting in literal rather than creative interpretations. Another ideation method is the checklist method (Hoshino, 2005). This method is easy to use because ideas can be generated in a short time, and unrelated concepts can be created using checklist items. However, because participants can freely select the items, it is difficult for design novices to develop new ideas. Therefore, they need to construct new conceptual elements from the theme before generating design proposals.

4 A PROPOSAL FOR AN IDEATION METHOD USING NOUN METAPHORS

4.1 Using Metaphorical Expressions

Because design novices have limited access to information, their ideas during divergent thinking tend to remain at the level of literal interpretation. To address this problem, design proposals using metaphorical expressions have been studied. Metaphors have also been shown to be effective in programming education for beginners (Mayer, 1978), suggesting their usefulness in learning. However, design proposals based on metaphors are often easy to understand because they remain close to literal interpretation, resulting in limited novelty. Therefore, this study focuses on the use of metaphors beyond literal expressions. According to Gibbs (2008), a metaphor such as "A is B" indirectly implies that "A is similar to B in some respect." In other words, a metaphor is interpreted as an indirect comparison based on analogy and similarity. By intentionally generating such metaphorical expressions, new ideas can be produced. In this study, novelty is defined as the degree to which an idea is new, original, and unfamiliar.

4.2 Leap Association Method(LAM)

During divergent thinking, it is difficult for design novices to break away from existing conceptual frameworks. Therefore, increasing conceptual distance is important. Isono and Takahashi (2023) reported that the novelty of ideas generated through conceptual combination is influenced by the differences and relationships between concepts. However, design novices find it difficult to increase conceptual distance abruptly. LAM instructs users to make conceptual leaps from a given theme by repeatedly generating metaphors consisting of the theme and a noun metaphor, thereby producing difficult-to-interpret metaphorical expressions. As a result, new conceptual elements are constructed from the theme. The final associated word is then used as a cue to generate a new design proposal (Figure 1). By increasing the conceptual distance from the original theme through repeated metaphorical associations, LAM enables the generation of novel design proposals. Therefore, this study hypothesizes that design novices can construct new conceptual elements and generate novel design proposals by repeatedly associating and using noun metaphors.

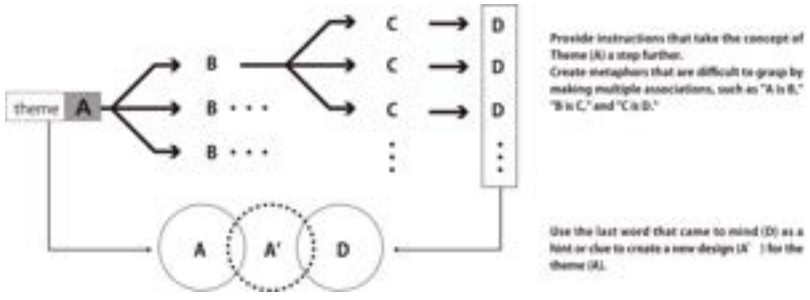


Figure 1. Schematic Diagram of LAM

5 RESEARCH METHODS

This study consists of two experiments. Experiment 1 examines the effectiveness of the LAM ideation process. Design students generated design proposals using LAM to determine

whether the method produces novel design proposals. The proposals were evaluated by design instructors, and conceptual distance was calculated from the association process. The relationship between novelty and conceptual distance was then examined. Experiment 2 compares LAM with the checklist method. In LAM, new conceptual elements are constructed from the theme before generating a design proposal. In contrast, in the checklist method, the timing of constructing new conceptual elements depends on the participant. Therefore, the relationship between the timing of concept construction and novelty was examined. In addition, differences in expert evaluations and participants' evaluations of the two ideation methods were investigated. Based on the results of these two experiments, the effectiveness of LAM as an ideation method for design novices was evaluated.

6 EXPERIMENT 1

6.1 Methods for Experiment 1

6.1.1 Word Generation and Design Proposals

Eighteen design students participated in the experiment. They were presented with Theme A (desk lamp) and instructed to think of a word describing another object that they believed had the same characteristics as the theme. They then wrote a syntagmatic metaphor in the form of "A is B" (Yamanashi, 2007). Based on the three-level structure of prototype theory (Rosch et al., 1976), the association task was repeated three times. Participants generated 21 words in total (Word B: 3, Word C: 9, and Word D: 9). Including the theme, the four-stage association process gradually increased the conceptual distance from the original theme. The associations were expressed as noun metaphors, in which both the theme and the metaphorical term were nouns and could be explicitly identified. After completing the word-generation task, participants exchanged worksheets. They then compared Theme A with Word D and selected one pair that they considered to be as conceptually distant from each other as possible (Figure 2). The selected Word D was used as a clue for generating a design proposal. Participants expressed their proposals as idea sketches, and drawing quality was not evaluated. To facilitate idea generation, Words B and C could also be used. Each worksheet included the proposal title, proposal characteristics, Theme A, and the selected word. Before the experiment, participants completed a practice session covering the entire procedure from word generation to design proposal.

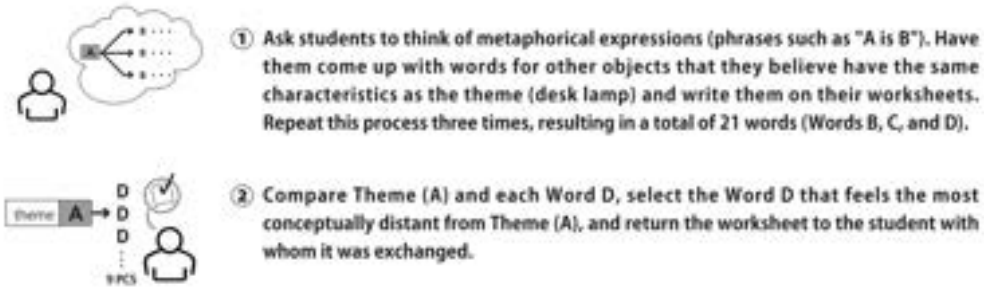


Figure 2. Steps for Writing a Sentence

6.1.2 Evaluation of Design Proposals

After completing the design proposal, participants self-evaluated the clarity of the ideation method using a five-point scale (1 = very unclear, 5 = very clear). They were then interviewed about the process of generating the associated words. Based on the interview results, the conceptual distance between associated words was classified into three levels (1 = direct, 3 = indirect). The classification was based on the Japanese Classification Vocabulary List (National Institute for Japanese Language and Linguistics, 2004) (Table 1). The scores were summed to obtain the total conceptual-distance score for each proposal. The idea sketches were evaluated by eight design instructors for novelty using a five-point scale (1 = not at all novel, 5 = very novel) based on the evaluation method proposed by Inoue (2005). During the evaluation, the instructors were provided with the idea sketch, proposal title, and proposal characteristics.

Table 1. Point Allocation Based on the Classification Vocabulary List

Evaluation Criteria	Differences in terminology, but between classification categories	Differences in classification categories, but between subcategories	Differences in departments
Scoring	1	2	3
Example	"Stove" "Induction Cooktop"	"Stove" "Griddle"	"Stove" "Sun"
	"Fishing Rod" "Whip"	"Fishing Rod" "Shower"	"Fishing Rod" "Giraffe"
	"Sun" "Moon"	"Sun" "Sunburn"	"Sun" "Fireworks"

6.2 Results of Experiment 1

6.2.1 Evaluation Results of Design Proposals

The experiment required 1.0–1.5 hours. A pilot study was conducted in advance to determine the number of associated words at each level. The mean self-evaluation score for the clarity of the ideation method was 2.9. Half of the participants selected "somewhat unclear" or "very unclear." For the expert evaluation, responses were obtained from eight design instructors. The mean novelty score of the design proposals was 3.4, with a maximum of 4.0 and a minimum of 2.6. Thirteen design proposals received a mean score above 3.0, indicating that more than half of the participants generated novel design proposals.

6.2.2 Results of Conceptual Distance Calculations Using a Classification Thesaurus

The mean conceptual-distance score increased slightly with each association (Table 2). For design proposals with a mean novelty score below 3.0, the mean conceptual-distance scores of the second and third associations were nearly the same. In contrast, for proposals with a mean novelty score above 3.0, the conceptual-distance score increased from the second to the third association. Thus, design proposals with higher novelty ratings showed greater conceptual distance at each stage of the association process. The total conceptual-distance scores for each design proposal showed the same tendency (Table 3). The correlation between the total conceptual-distance score and the mean expert novelty score showed a weak correlation ($R^2 = 0.40$) (Figure 3).

Table 2. An example of a table

participant	A	B	Background and Scores for AB		C	Background and Scores for BC		D	Background and Scores for CD	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	Desk lamp	Plates	The plate I have at home is the same color as my desk lamp	2	Desks	Imagine what's there	2	People	Imagine the legs of a desk	3
					Inner tubes	The same shape	1	Water sliders	Imagine a pool	2
					Rice	Inspired by a rice bowl	2	Perfume	Imagine the rice being cooked	2
		Sun	A bright image	3	Moon	Types of Planets	1	Idol	A sparkling image	3
					Fireworks	A glowing image	3	Fireflies	A glowing image	3
					Kotatsu	A warm and cozy feel	3	Socks	Imagine putting your foot in	2
		Textbook	The image of studying	3	Chocolate	Imagine a square shape	3	Winter	Buy and eat during the winter season	3
					Air	A breeze blows when you turn the page	3	Cigarettes	Imagine a gas	3
					Bacon	A thick image	3	Skin	The image of something burning	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Average score		2.0			2.2			2.3		

Table 3. The Relationship Between Conceptual Distance Scores and Expert Evaluations

Participant		Average Expert Rating	Total score for 1st round (between AB) (out of 9 points)	Total score for 2nd round (between BC) (out of 27 points)	Total score for 3rd round (between CD) (out of 27 points)	Total score (out of 63 points)
1		3.1	5	13	20	38
2		3.8	4	17	24	45
3		3.9	4	15	24	43
4		2.8	4	20	19	43
5		3.9	6	22	21	49
6		3.9	6	20	21	47
7		3.6	7	19	26	52
8		3.6	8	21	24	53
9		3.8	4	20	24	48
10		4.0	7	27	21	55
11		3.4	5	23	23	51
12		4.0	8	23	20	51
13		2.8	8	19	16	43
14		2.6	4	14	17	35
15		3.3	9	17	17	43
16		3.5	8	26	19	53
17		2.6	6	22	17	45
18		2.8	6	20	18	44
Avg.	All Designs	3.4	6.1	19.9	20.6	46.6
	Design proposals exceeding 3 points	3.7	6.3	20.8	22.2	49.2
	Design proposals falling below 3 points	2.7	5.5	18.0	17.8	41.3

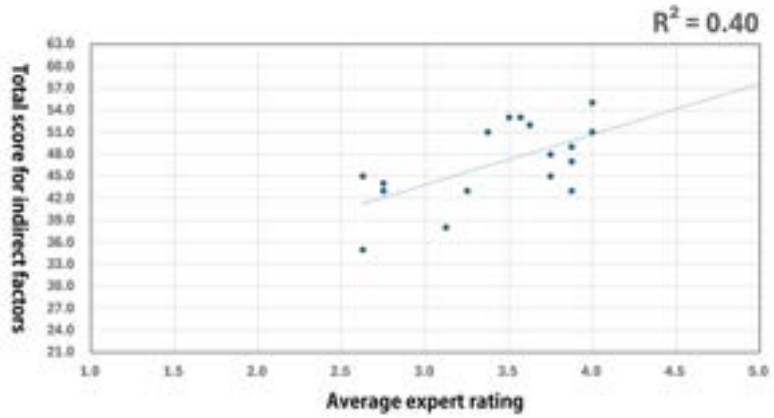


Figure 3. Correlation Chart Between Conceptual Distance and Novelty

6.3 Discussion of Experiment 1

Participants increased the conceptual distance from the original theme through repeated associations. As the associated words became more distant from the theme, novelty was generated through unexpected associations, which may explain the weak correlation between conceptual distance and the expert novelty evaluation. The limited selection of associated words may also have restricted the novelty of the design proposals. Eight participants did not use Word D directly. Four generated proposals using words similar to or derived from Word D (Figure 4, left), whereas the remaining four used Word B or Word C (Figure 4, right). These proposals did not fully utilize the conceptually distant word. However, when Word D was difficult to use, intermediate words functioned as alternative cues, suggesting a secondary effect of LAM. This experiment provides insight into the association process and forced association. However, because only LAM was evaluated, its effectiveness could not be fully verified. Therefore, Experiment 2 compares LAM with the checklist method.

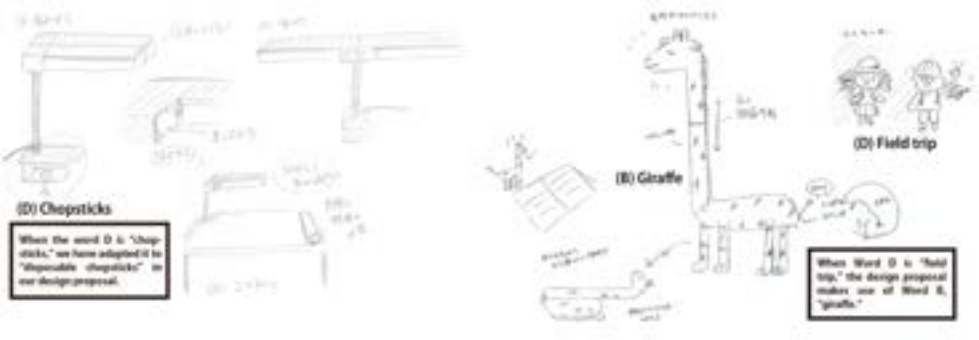


Figure 4. Examples of Design Proposals

7 EXPERIMENT 2

7.1 Methods for Experiment 2

Twenty design students generated design proposals using both LAM and the checklist method. The theme for LAM was a water bottle, whereas the theme for the checklist method was earphones. The checklist method was based on Osborn's checklist. Because the original checklist uses formal expressions, the items were rewritten in colloquial language with reference to a design textbook (Kawamen & Suga, 2003) (Table 4). Participants could use multiple checklist items. To shorten the experiment, the number of associated words was reduced while maintaining three association stages (Word B: 3, Word C: 6, Word D: 6). Participants were divided into two groups. One group used LAM followed by the checklist method, whereas the other group used the reverse order. The two sessions were conducted at least two days apart. After completing the design proposals, participants evaluated the clarity of each ideation method using a five-point scale and provided written comments. They then compared their two design proposals and selected which method was easier for generating design proposals and which proposal was more novel. The design proposals were evaluated by eight design instructors. The proposals were randomly assigned to the

evaluators using the RAND and MOD functions. Statistical analyses were conducted to examine differences between the ideation methods and among the evaluators.

Table 4. The Relationship Between Conceptual Distance Scores and Expert Evaluations

No.	1	2	3	4	5	6	7	8	9
Literary	(Put to) Other uses	Adapt	Modify	Magnify	Minify	Substitute	Rearrang	Reverse	Combine
Colloquial	Are there any other uses for this?	Can this be applied to other areas?	Can this be modified?	What if we make it bigger?	What if we make it smaller?	What if I use a substitute?	What if I swap them out?	What if I reverse it?	What if I combine them?

7.2 Results of Experiment 2

7.2.1 Participants' Self-Evaluation

The mean score for the clarity of the ideation method was 4.0 for LAM and 3.8 for the checklist method. A statistical test showed no significant difference between the two methods ($\alpha = 0.05$, two-tailed, $p = 0.65$). Therefore, the clarity of the two ideation methods was considered comparable. When participants compared the two methods, eight selected LAM as being easier for generating design proposals, whereas ten selected the checklist method. However, when asked which design proposal was more novel, 13 of the 20 participants selected the proposal generated using LAM (Table 5).

7.2.2 Expert Evaluation

The mean expert novelty score was 3.2 for LAM and 2.9 for the checklist method. Although the mean scores differed between the first and second evaluator groups, both groups assigned higher novelty scores to LAM than to the checklist method. A statistical test showed no significant difference between the two ideation methods ($\alpha = 0.05$, two-tailed, $p = 0.32$). Among the eight design instructors, five gave higher novelty ratings to LAM. Of these, two showed a statistically significant preference for LAM (Table 6).

Table 5. Participant Self-Assessment Results

Participant		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Degree of Clarity in Ideation Methods	LAM	5	4	5	2	5	3	3	2	2	4	5	4	5	4	3	5	4	5	5	4	4.0
	Chekclist Method	4	5	4	4	4	4	2	4	4	5	4	2	3	2	4	4	5	4	4	3	3.8
Which method made it easier to propose designs?		LAM	Neither	CHK	Neither	LAM	CHK	LAM	CHK	CHK	CHK	LAM	CHK	CHK	LAM	CHK	LAM	CHK	LAM	CHK	LAM	Significant difference (t-test, p < 5%, two-tailed)
Which proposal was considered a new idea?		CHK	CHK	LAM	LAM	LAM	LAM	LAM	LAM	LAM	LAM	Neither	LAM	LAM	LAM	LAM	Neither	CHK	Neither	CHK	LAM	No

Table 6. Expert Assessment Results

	1st half				1st half Avg.	2nd half				2nd half Avg.	Avg.	Significant difference (t-test, Two-tailed test (α = 5%, p < 0.05))
Design instructors	1	2	3	4		5	6	7	8			
LAM	4.6	3.5	3.0	3.0	3.6	3.5	2.9	2.3	2.4	2.8	3.2	No
Chekclist method	3.9	3.2	3.6	2.7	3.4	3.2	2.8	1.5	2.6	2.5	2.9	
Significant difference (t-test: One-tailed test (5% significance level))	Yes	No	No	No	—	No	No	Yes	No	—	—	—

7.3 Discussion of Experiment 2

The mean expert novelty score was higher for LAM than for the checklist method, although statistically significant differences were found only in a limited number of cases. This may be

because the evaluation criteria were not explicitly presented to the evaluators. Using paired comparisons instead of rating scales may improve the evaluation process. Participants, however, regarded the LAM proposals as more novel. Comments on the checklist method indicated that its proposals tended to be more practical and that selecting checklist items required more time. Most participants expressing these opinions preferred the LAM proposals (Table 7). Although participants considered the checklist method easy to use, they evaluated its novelty less favorably. These evaluations may have been influenced by personal preferences. However, because LAM requires participants to construct new conceptual elements from the theme, the results suggest its potential effectiveness.

Table 7. Comparison of self-assessments on clarity, comments, and originality (partial)

Participant	Checklist method		LAM		Which is the newer proposal: LAM or the Checklist method?
	Level of clarity	Comments	Level of clarity	Comments	
1	4	I couldn't come up with any ideas. I found it difficult to think outside the box.	5	By proposing ideas from things that seemed completely unrelated, I was able to broaden my way of thinking.	Checklist method
4	4	It was easy to get started, and ideas came to me easily, but the result ended up being rather cliché.	2	It took me a while to come up with any ideas.	LAM
5	4	It felt like it took a long time for ideas to come to me.	5	I felt that designing by relating it to something else, as I did this time, made it easier to come up with ideas.	LAM
6	4	It felt like the process went smoothly. However, it was hard to come up with truly original ideas.	3	It was somewhat forced, but it went smoothly; however, I think the chosen words might make it difficult to express the concept.	LAM
8	4	I was pressed for time, so it was difficult to think deeply about things.	3	It was interesting how, when associating words, I tried to break down the elements of each word to find its essence.	LAM
10	5	No matter what I did, I kept coming up with only practical ideas.	5	I was worried I wouldn't be able to come up with any ideas because I strayed from the theme, but the process itself was interesting.	LAM
18	5	I wasn't sure which one to choose, so it felt like it took a little longer than expected.	5	It was difficult, but I was able to think of ideas from perspectives I wouldn't normally notice.	Neither
20	3	I found it difficult to come up with ideas.	3	I feel like the initial metaphor really stuck with me.	LAM

8 CONCLUSION

This study evaluated the effectiveness of LAM for generating novel design proposals by design novices through two experiments. In Experiment 1, design proposals that received higher expert novelty scores also showed higher conceptual-distance scores. In addition, the conceptual-distance score increased with each association, indicating that participants gradually increased the conceptual distance from the original theme while generating design proposals. In Experiment 2, LAM was compared with the checklist method. Although a statistically significant difference in expert novelty evaluation was confirmed only in a limited number of cases, the clarity of the ideation method was comparable between the two methods. Moreover, more than half of the participants evaluated the design proposals generated using LAM as more novel. These results provide partial support for the effectiveness of LAM. LAM uses the theme and the final associated word to generate design proposals. However, some participants used intermediate associated words or did not use the final associated word directly, making design proposal generation difficult in some cases. To improve the method, multiple conceptually distant words should be prepared so that participants can select alternative cues when necessary. LAM is intended to support divergent thinking. Therefore, the resulting design proposals may have limited usefulness and feasibility, indicating that further improvement of the method is required. Although this study focused on novelty, particularly unexpectedness and originality, future studies should also examine other design qualities. Because LAM depends heavily on language, future research should also consider linguistic and cultural contexts when developing the method.

9 FUTURE WORK

Future research aims to generalize LAM by establishing it as an ideation method for practical design exercises and educational settings. The method will be improved by comparing the use and non-use of conceptually distant words and by examining how associated words should be used during the ideation process.

REFERENCES

Kumar, V. (2015). 101 Design Methods (Watabe, N., Trans.). Eiji Press, pp.14-16.

Murayama, Y. & Sato, H. (2025). A Proposal for a Method of Generating Ideas for Design Novices. Japanese Society for the Science of Design the 72nd Spring Research and Presentation Conference, A6-07. https://doi.org/10.11247/jssd.72.0_78

Watanabe, M., Mori, N., & Sugiyama, K. (1984). Developing a System to Get Suitable Formative Elements by Using of Fuzzy Inference. Bulletin of Japanese Society for the Science of Design, 88, pp.145–152. https://doi.org/10.11247/jssdj.1992.145_1

Hoshino, M. (2005). Hassouhounyuumon(Introduction to Creative Thinking). Nikkei Business Publications, Inc. , pp.128-134.

Tanaka, T. & Noguchi, N. (2004). Denial Expression Method for Creative Concept Generation. Bulletin of Japanese Society for the Science of Design, 51(3), pp.39-46. https://doi.org/10.11247/jssdj.51.39_3

Mayer, R. (1978). Advance Organizers That Compensate for the Organization of Text. Journal of Educational Psychology, 70(6), pp.880–886. https://doi.org/10.1037/0022-0663.70.6.880

Gibbs, R. W., Jr. (2008). The Poetics of Mind (K. Yagi, Trans.). Kenkyusha Co., Ltd., pp.227-234.

Isono, M., & Takahashi, K. (2023). The Use of Conceptual Combination and Analogy and Their Relationship in Creative Idea Development. Bulletin of Japanese Society for the Science of Design, 69(4), pp.29–38. https://doi.org/10.11247/jssdj.69.4_29

Yamanashi, M. (2007). Hiyu to Rikai(Metaphors and Understanding). University of Tokyo Press., pp.15-20.

Rosch, E., Mervis, C., & Wayne, G. D. (1976). Basic Objects in Natural Categories. Cognitive Psychology, 8(3), pp.382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Inoue, K. (Ed.). (2005). Design and Kansei. KAIBUNDO PUBLISHING CO., LTD., pp.179-181

National institute for japanese language and linguistics (Ed.). (2004). Classification Thesaurus. Dainippon Tosho Co.,Ltd.

Kawamen, E., & Suga, M. (2003). Souzoutekikougakusekkei No Houhou(Methods of Creative Engineering Design). Yokendo Ltd. Publishers, pp.168-169.

PREDICTION OF CLOTHING COMFORT DURING SWEATING USING MULTIMODAL EXPLAINABLE MACHINE LEARNING

Shun MOROOKA[a], Hiroaki YOSHIDA[a] , Masayoshi KAMIJO[b]

[a] Shinshu University, Ueda, Nagano, 386-0018, Japan, kamijo@shinshu-u.ac.jp

ABSTRACT

Clothing comfort during sweating is influenced by interactions among psychological perceptions, physiological responses, and textile material properties. This study investigated whether interpretable machine-learning models could predict subjective clothing comfort and identify the factors associated with it. Five polyester sportswear garments differing in knit structure, thickness, mass per unit area, and moisture-management finish were evaluated under controlled environment conditions (30 °C and 65%RH). Ten participants completed repeated exercise-rest cycles. Subjective and physiological responses were recorded at five time points for each garment, while fabric mechanical and water-absorption properties were measured separately. A total of 250 observations (10 participants x 5 garments x 5 time points) were analyzed as a single regression task using a generalized linear model, decision tree, random forest, and support vector machine. The support vector machine achieved the numerically lowest RMSE (0.560 ± 0.060). Ease of movement and perceived surface fineness were consistently identified as important predictors across models. These findings demonstrate the potential of multimodal machine learning for predicting clothing comfort during sweating and identifying the factors associated with subjective comfort.

Keywords: Clothing Comfort, Nonlinear Modeling, Sportswear, Physiological Measurement, Textile Material Properties

1 INTRODUCTION

Clothing comfort is a multidimensional perception that arises from interactions among the physical properties of clothing, the wearer's physiological responses, psychological perceptions, and the environmental conditions (Li, 2001). In sportswear and other functional clothing, these interactions change dynamically during physical activity and sweating, and sensations such as wetness, stickiness, thermal sensation, and tactile sensation become important determinants of comfort (Raccuglia et al., 2018). Because these factors interact in complex and potentially nonlinear ways, evaluations based solely on individual material properties or conventional linear analyses may be insufficient to explain overall clothing comfort.

Previous studies have demonstrated that overall perceived clothing comfort during exercise can be predicted from multiple subjective sensory evaluations using artificial neural network (ANN) models (Wong et al., 2003). Integrating psychological perceptions with physiological responses has also been shown to improve the prediction of clothing comfort compared with models based on either type of information alone (Karasawa et al., 2022). Collectively, these findings suggest that clothing comfort should be regarded as a multimodal and potentially nonlinear outcome. Machine-learning approaches are therefore well suited to integrating heterogeneous data and capturing complex relationships among variables that may not be adequately represented by conventional single-domain or linear analytical methods. However, previous studies have primarily focused on subjective evaluations or combinations of psychological and physiological responses, whereas the simultaneous integration of psychological, physiological, and textile material variables during sweating has received limited attention. Accordingly, this study compared four regression algorithms using multimodal data comprising psychological evaluations, physiological responses, and fabric material properties. The objectives were to evaluate the predictive performance of these models and to identify the variables consistently associated with perceived clothing comfort during repeated exercise-rest cycles.

2 EXPERIMENTS

A wear trial was conducted under controlled environment conditions (30 °C and 65% RH) using five 100% polyester knitted sportswear garments differing in knit structure, thickness, mass per unit area, and water-absorbing or water-repellent finish. Mechanical and surface properties were evaluated using the Kawabata Evaluation System (KES), and water absorbency was measured according to JIS L 1907 by the drop and Byreck methods.

Ten healthy male university students aged 22-24 years completed a protocol consisting of a 10-min seated rest followed by four cycles of 10-min cycling at 50 W and 60 rpm interspersed with 5-min rest periods. Subjective ratings and physiological responses were recorded at five predefined time points for each garment. Physiological measurements included skin and core temperatures, mean body temperature, clothing microclimate humidity, sweat rate, electrocardiographic (ECG) signals, and heart-rate-variability (HRV) indices. Semantic differential scales were used to assess perceived wetness, stickiness, slipperiness, stiffness, roughness, softness, ease of movement, and overall clothing comfort. Body mass and garment mass were measured before and after the experimental protocol.

Overall clothing comfort at each measurement point was used as the only target variable and ranged from -3 (uncomfortable) to +3 (comfortable); therefore, it was excluded from the predictor variables. Each measurement point was treated as a separate observation for predictive modeling. Time point and elapsed time were not included as predictors because the objective was to estimate comfort from the concurrent psychological, physiological, and material states at different stages of sweating rather than to model temporal transitions.

The dataset consisted of 250 observations (10 participants × 5 garments × 5 measurement points). Four regression models-- a generalized linear model (GLM), decision tree, random forest, and support vector machine (SVM)—were developed using the same standardized predictor set. The observations were randomly divided into training (60%) and validation (40%) sets. The

procedure was repeated using seven randam train – validation splits ; the highest and lowest RMSE values were discarded, and the mean ± SD of the remaining five runs was reported. Model-specific feature importance was interpreted within each algorithm, and the consistency of highly ranked variables across models was examined.

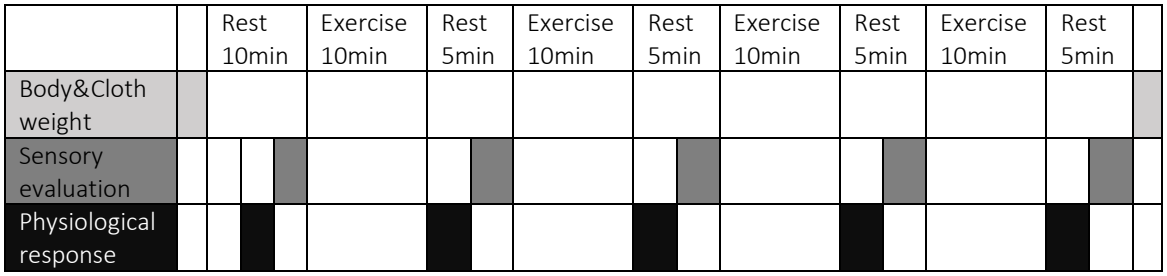


Figure 1. Experimental Procedure

3 RESULTS & DISCUSSION

Table 1 summarizes the prediction performance of models and the representative important predictors identified by each algorithm. The SVM achieved the lowest mean RMSE of 0.560 ± 0.060, followed by the GLM (0.586 ± 0.092) and random forest (0.598 ± 0.111). The decision tree exhibited the highest RMSE (0.690 ± 0.126). Given that the target variable was evaluated on a seven-point scale ranging from -3(uncomfortable) to +3(comfortable), these RMSE values correspond to an average prediction error of less than one scale point. The results indicate that subjective clothing comfort can be predicted with acceptable accuracy using multimodal psychological, physiological, and material information. However, the differences in prediction performance among the GLM, SVM, and random forest were relatively small, suggesting that all three approaches showed comparable predictive capabilities for this dataset.

Table 1. Prediction performance and representative important variables.

Model	RMSE	Representative high-ranked variables
GLM	0.586 ± 0.092	Hard to move - Easy to move; (Body) Cold - Hot; Coarse - Fine; Mean body temperature ; Top Absorption
Decision tree	0.690 ± 0.126	Coarse - Fine; Hard to move - Easy to move; Arm sweat rate; Rough - Smooth; Hard - Soft
Random forest	0.598 ± 0.111	Heavy - Light; Hard to move - Easy to move; Coarse - Fine; Hard - Soft; Compression
SVM	0.560 ± 0.060	Coarse - Fine; Hard to move - Easy to move; Heavy - Light; Loose - Tight; CVR-R

The variables identified as important predictors differed among models, and should not be interpreted as direct evidence of causality. Rather, they indicate that each model captured the associations between clothing comfort and the measured variables in different ways. Nevertheless, the semantic differential scales *Hard to move - Easy to move* and *Coarse - Fine* were

consistently ranked among the most important predictors across the models. These variables can therefore be considered robust indicators of clothing comfort under sweating conditions. Ease of movement is likely to reflect the combined effects of garment fit, fabric deformation, and mechanical resistance during body movement, whereas perceived surface fineness may be associated with skin-contact sensations, fabric surface characteristics, and adhesion between the skin and moisture-laden clothing.

Physiological and material variables were also identified as important predictors in several models. Mean body temperature, arm sweat rate, and CVR-R were among the variables associated with clothing comfort prediction, suggesting that thermal and autonomic responses may contribute to perceived comfort under hot and humid conditions. Top Absorption and Compression were likewise identified as important predictors, indicating that moisture absorption and mechanical properties may influence perceived comfort indirectly through their effects on the clothing microclimate and tactile sensation. These findings highlight the value of integrating psychological, physiological, and material information when predicting clothing comfort. However, because the dataset comprised repeated measurements from the same participants and garment samples, these findings should be interpreted as exploratory rather than conclusive. Future studies should evaluate generalizability of the models using participant-wise or sample-wise validation and investigate the temporal dynamics of clothing comfort perception during repeated exercise-rest cycles.

4 CONCLUSIONS

This study investigated whether subjective clothing comfort during sweating could be predicted using psychological evaluations, physiological responses, and fabric material properties. Among the evaluated models, the SVM achieved the lowest RMSE, whereas the GLM and random forest showed comparable predictive performance. Ease of movement and perceived surface fineness were consistently identified as important predictors, suggesting that these sensations are robust indicators of clothing comfort. Physiological and material variables were also identified as important predictors across several models, highlighting the value of multimodal machine learning for identifying factors associated with clothing comfort. Overall, the findings demonstrate the potential of interpretable multimodal machine learning as a framework for predicting clothing comfort while providing insight into the psychological, physiological, and material factors associated with comfort perception. Future studies should validate the proposed framework using independent participant- and sample-wise datasets.

REFERENCES

Li, Y. (2001). The science of clothing comfort. *Textile Progress, 31*(1–2), 1–135.

Raccuglia, M., Sales, B., Heyde, C., Havenith, G., & Hodder, S. (2018). Clothing comfort during physical exercise—Determining the critical factors. *Applied Ergonomics, 73*, 33–41.

Wong, A. S. W., Li, Y., Yeung, P. K. W., & Lee, P. W. H. (2003). Neural network predictions of human psychological perceptions of clothing sensory comfort. *Textile Research Journal, 73*(1), 31–37.

Karasawa, Y., Uemae, M., Yoshida, H., & Kamijo, M. (2022). Prediction of clothing comfort sensation of an undershirt using artificial neural networks with psychophysiological responses as input data. *Textile Research Journal, 92*(3–4), 330–345.

PROPOSED DESIGN FRAMEWORK FOR DIGITAL PIANO INTERFACES: A USER-CENTERED APPROACH

Meng-Cong Zheng[1a]; Ching-I Chen[b]; Li-Jen Wang[c], Fusako Ishimura[d], Yasuo Shiozawa[e], Wen-Hsien Kuo[f], Ai-Ruo Chang[g], Kuan-Yu Chen[h], Chia-Ling Zeng[i], Hsuan-Wu Sung[j], Yen-Hsiang Chang[k]

[a] Department of Industrial Design, National Taipei University of Technology, Taiwan, zmcdesign@gmail.com
[b] Department of Industrial Design, National Taipei University of Technology, Taiwan, zoeychen116@gmail.com
[c] College of Design, National Taipei University of Technology,Taiwan, lilijen1013@gmail.com
[d] Yamaha Corporation, Japan, fusako.ishimura@music.yamaha.com
[e] Yamaha Corporation, Japan, yasuo.shiozawa@music.yamaha.com
[f] Department of Industrial Design, National Taipei University of Technology,Taiwan, hsienxx@gmail.com
[g] Department of Industrial Design, National Taipei University of Technology,Taiwan, t113588009@ntut.org.tw
[h] Department of Industrial Design, National Taipei University of Technology,Taiwan, m901372@gmail.com
[I] Department of Industrial Design, National Taipei University of Technology,Taiwan, gretazeng915@gmail.com
[J] Department of Industrial Design, National Taipei University of Technology,Taiwan, dino2413@gmail.com
[K] Department of Industrial Design, National Taipei University of Technology,Taiwan, ea167077@gmail.com

ABSTRACT

With advances in music technology, digital pianos now provide high-quality sound simulation and increasingly diverse functions. However, interface design faces a growing challenge in balancing feature expansion with ease of use. Fixed physical buttons are limited in supporting complex parameter adjustments and may increase extraneous cognitive load, while touchscreen interfaces often lack the tactile feedback needed for embodied interaction. This study aims to propose a user-centered design framework to address these challenges.

A mixed-methods approach was conducted in two phases. In Phase 1, in-depth interviews with 10 experts were used to identify key operational preferences and functional layout principles. In Phase 2, a questionnaire survey was administered to 390 digital piano users. Cluster analysis

[1] Corresponding author. 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608 Taiwan, zmcdesign@gmail.com

identified three user groups: Advanced Experts, Intermediate Practitioners, and Novice Entry Users. Results showed that 71.0% of users preferred physical buttons during individual practice, with Intermediate Practitioners demonstrating particularly high reliance on physical controls (96.7%). However, a mismatch was observed between current interface designs and actual usage patterns. For example, although the Transpose function is often prominently positioned on control panels, 63.6% of users reported rarely or never using it, suggesting inefficient use of physical interface space. Additionally, simplified menus (58.5%) and one-touch shortcuts (55.4%) were identified as key factors for improving operational efficiency.

Based on these findings, this study proposes a three-layer design framework consisting of a Functional Layer, Interaction Layer, and Emotional Layer. The framework emphasizes reorganizing functions according to usage frequency and integrating tactile and visual feedback to reduce cognitive load and improve usability across different user groups.

Keywords: *Expert tacit knowledge, Kansei engineering, Visual compensation, Embodied cognition, Cluster analysis*

1 INTRODUCTION

With rapid advancements in music technology, Digital Musical Instruments (DMIs) have evolved from simple sound imitation into complex systems with high synthesis capabilities and precise parameter control. Unlike traditional acoustic instruments, the core characteristic of digital instruments lies in the physical separation of the "control interface" and the "sound generation engine" (Jensenius, 2022). While this separation offers design flexibility, it also reduces the physical coupling between the player and the instrument, making intuitive control a challenge (Hunt & Kirk, 2000; Wessel & Wright, 2002). Among numerous digital instruments, digital pianos have become the preferred choice for modern music education and home practice due to their simulation of traditional piano touch, silent practice capabilities, maintenance-free tuning, and diverse sound libraries. However, this multifunctionality has led to extreme complexity in the operation interface; while traditional piano keyboards provide excellent mechanical tactile feedback, their static physical layout struggles to correspond with increasingly complex digital functions and parameter adjustments (Jensenius, 2022).

To resolve the conflict between functional expansion and limited interface space, existing digital pianos often adopt a multi-level menu operation logic. This forces users to navigate through layers of menus, leading to increased "extraneous cognitive load" (Sweller, 2011), or to adopt a "second screen" strategy using external Apps. Although touchscreen interfaces offer layout flexibility (Shaer & Hornecker, 2010), they lack physical tactile guidance and damping. This increases the need for visual confirmation, conflicting with the "embodied cognition" nature of performance that emphasizes physical perception and eyes-free operation (McPherson, 2012; Schiavio & Menin, 2013). Furthermore, from the perspective of Kansei Engineering, when visual expectations do not match actual tactile feedback, it affects the user's emotional experience and pleasure (Yanagisawa & Takatsuji, 2015).

The theoretical dilemmas mentioned above are concretely embodied in high-end flagship models like the Yamaha P-525. Although the P-525 possesses a high degree of acoustic simulation maturity with its GrandTouch-S keyboard mechanism and CFX/Bösendorfer binaural sampling technology (Yamaha, 2023), its traditional "fixed physical button" interface, used for many years, has gradually become a structural bottleneck limiting User Experience (UX) upgrades. This "one-size-fits-all" design lacks flexibility when facing heterogeneous needs (Norman, 2013) and creates significant cognitive conflicts among users of different professional levels. For beginners, the complex and unguided panel configuration easily causes excessive cognitive load, hindering learning outcomes (Sweller, 1994, 2011). For professional performers pursuing efficiency, fixed buttons force them to make adjustments to key parameters through tedious "Menu Diving." Community feedback and professional reviews indicate that this design seriously interferes with the immediacy of stage performance and the state of flow (Welter & Ren, 2025).

To break through this limitation, this study attempts to address the rigidity of physical button operations, enabling the interface to be flexibly adjusted according to the user's contextual needs. On one hand, it provides simplified navigational guidance for novices, on the other hand, it grants experts the authority to define shortcuts, aiming to effectively reconcile the contradiction between "beginners' need for simplified navigation" and "professional players' pursuit of operational efficiency" (Jorritsma et al., 2015). Specifically, this study proceeds in two stages: first, gaining an in-depth understanding of the current usage of the Yamaha P-525 and its digital applications; second, analyzing user pain points in operating the integrated physical interface and virtual menus through task-oriented experiments and questionnaires; and finally, based on literature analysis and experimental results, proposing and validating a design proposal for a new interface, aimed at providing concrete, User-Centered design suggestions for the upgrading of the digital instrument industry.

2 METHODS

This study employs a mixed methods approach, conducted in two phases. Phase one involves expert interviews to deeply explore the operational perspectives and interface concepts of professional users. Phase two uses a questionnaire survey to validate the needs and preferences of large-scale users. Finally, data from both phases are integrated to summarize the key factors affecting the digital piano experience and derive a design framework.

2.1 Phase One: Expert Interviews

2.1.1 Participants

This phase adopted purposive sampling to recruit 10 participants aged 18 or older with experience operating electronic keyboards or digital pianos. Additionally, all interviewees were required to have at least 3 years of teaching experience to ensure their professional familiarity with teaching and practical usage. Interviews were conducted either in-person or online.

2.1.2 Interview Content

The interview content consisted of two parts: oral response and practical operation. The oral part covered three major dimensions: operational preferences, practical teaching pain points,

and future outlooks, comprising 11 open-ended questions. The practical part introduced the concept of Participatory Design, utilizing the CANVA Whiteboards collaboration platform. Interviewees were tasked with two specific activities: first, to rank a provided list of Yamaha P-525 functions based on their usage frequency and explain the rationale behind their ranking; and second, to interact with a top-down view of the digital piano by using a mouse to drag and rearrange or add function buttons, as well as adjust the screen size according to their operational preferences. These exercises aimed to create an "Ideal P-525 Interface Layout" (Figure 1) and concretize the interviewees' imagination regarding function zoning and operation logic.

Regarding data analysis, interview transcripts were analyzed using thematic analysis. Potential core issues were identified through coding and categorization. Two researchers independently coded the data to identify implicit and meaningful themes and concepts within the interviews. Reliability of the coding was ensured through repeated reading and the constant comparative method. Cohen's kappa coefficient was calculated to determine the consistency between the two coders. For the visual layouts produced in Canva, the heat zones of configuration and hierarchical logic of functional blocks, as well as function usage frequency, were summarized to serve as the basis for the subsequent questionnaire design.

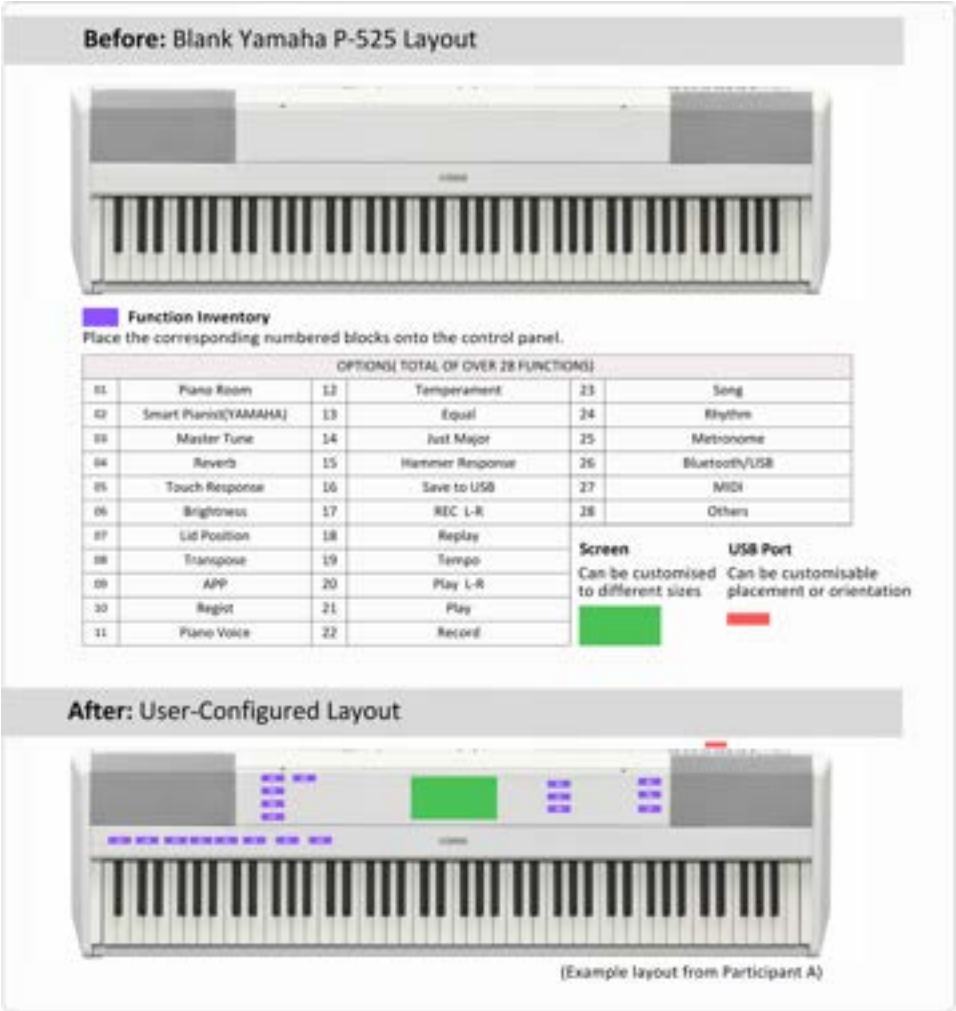


Figure 1. Schematic of the participatory design task using CANVA Whiteboards.

2.2 Phase Two: User Survey

This phase adopted an online questionnaire survey method, targeting social media forums and groups related to digital pianos as the main distribution channels. Participants were required to be 18 years or older and have experience operating digital pianos. To ensure the sample data possessed sufficient representativeness and reliability for subsequent statistical analysis, this study set corresponding sample size goals, ultimately collecting 390 valid questionnaires. The questionnaire items were based on the qualitative analysis results from the first phase of expert interviews, totaling 28 questions. The structure covered five core dimensions: identifying the participant's background and professional level through "Basic Information" and "Piano Learning History"; exploring "Operational Experience and Perception" to evaluate user satisfaction and cognitive load with existing interfaces; comparing user acceptance of physical buttons, touchscreens, and external Apps through the "Interface Type Preference" dimension; and finally collecting specific qualitative feedback through "Open-ended Suggestions".

2.2.1 Questionnaire Analysis

Quantitative data were analyzed using SPSS software. First, descriptive statistics were used to present the sample structure. Then, a Two-step cluster analysis was performed to group participants based on their learning history and skill indicators. Subsequently, one-way ANOVA was used to explore significant differences between different clusters in "Operational Experience" and "Interface Preference," thereby identifying the key needs of different groups to serve as the design basis for the interface framework.

3 RESULTS

3.1 Expert Perspectives

This study conducted in-depth interviews with ten experts. The interview data were subjected to thematic analysis by two researchers who did not participate in the interviews to ensure the objectivity and consistency of the coding process. The analysis results summarized five main themes, involving hardware and software design/application of digital pianos, external system connection, directions for improvement, and market positioning. Through interviews with digital piano teachers, the aim was to explore the actual usage scenarios and future development trends of digital pianos in the current market.

This study used thematic analysis for data processing. To ensure coding reliability, two researchers independently judged 709 coding units. The Cohen’s Kappa consistency test showed a coefficient of $\kappa = 0.823$, $SE = 0.019$, $\chi^2(1) = 37.489$, $p < .001$. According to the standards of Landis and Koch (1977), this indicates an almost perfect agreement between raters.

The interview results show that in practical application, digital piano core functions are the metronome, tone adjustment, recording, and automatic accompaniment. Especially in teaching scenarios, teachers often use these functions to improve student learning outcomes and correct performances. Regarding interface design and operational preferences, although Bluetooth and Apps (such as Smart Pianist) provide expandability and recording convenience, users generally still prefer intuitive physical button operations; models that rely excessively on App control (such

as the CSP series) present a higher barrier to operation. Regarding hardware specifications, interviewees emphasized screen visibility (e.g., font size for vision protection) and the necessity of USB/Bluetooth connectivity; screens that are too small may reduce the willingness of older adults or students to use them. regarding market positioning, although digital pianos are seen as an alternative to traditional pianos due to budget and space considerations, they still face competition from portable keyboard instruments. For future development, interviewees suggested continuing to optimize the authenticity of touch and tone, and providing Traditional Chinese interfaces and complete after-sales support to meet the needs of users ranging from beginners to professional musicians.

3.2 User Survey

3.2.1 Participant Demographics

This study collected 390 valid questionnaires. The participant profile shows the diversity of digital piano users. In terms of demographics, the gender ratio was fairly balanced (Female 53.8%, Male 46.2%). The age group was mainly concentrated in the middle-aged group of 36 to 55 years old, accounting for 72.9% of the total sample. Regarding identity, the vast majority of participants were office workers or freelancers (83.8%), while music majors or professional performers accounted for a very low percentage. Notably, in self-assessed playing skill, over 60% of participants (62.8%) considered themselves "Beginners," indicating that digital pianos are primarily used as learning or leisure tools for non-professionals. regarding brand ownership, Yamaha is the dominant market brand (54.6%), followed by Casio (13.3%) and Kawai (7.2%), while 28.2% of participants do not currently own a digital piano. regarding usage scenarios, the vast majority of users (75.9%) use digital pianos for "Personal Practice," followed by "Accompanying others' practice" (21.3%) and "Accompaniment/Ensemble" (16.4%), showing that personalized practice remains the core application scenario for this type of instrument.

3.2.2 Interface Preferences and Operational Challenges

Analysis of interface preferences across different usage scenarios shows that physical buttons remain the most relied-upon interaction mode in most situations. Particularly in the "Personal Practice" scenario, as many as 71.0% of users prefer using physical buttons, far higher than touch buttons (30.5%) and display screens (20.0%). However, as task complexity increases, such as in "Music Production" or "Accompaniment/Ensemble" scenarios, user acceptance of touch buttons, screen interfaces, and external Apps increases significantly. In terms of App usage behavior, Yamaha's Smart Pianist is the application with the highest penetration rate; about half of the users have experience using it, mainly for assisted practice and accompaniment. (Table X_Q 15)

The in-depth analysis of usage behavior further reveals a phenomenon of high idle rates for built-in functions, highlighting an asymmetry between "function supply" and "actual demand." Data shows that even for functions marketed as core features of digital pianos, their frequent usage rate (sum of "Often" and "Very often") is generally low. For example, "Transpose" and "Recording" functions, often placed prominently on the panel, have as many as 63.6% and 55.8% of users respectively stating they "Never" or "Rarely" use them. Even for relatively popular functions like "Rhythm" and "Voice/Change Voice," the proportion of high-frequency usage (weekly or more) is only between 25% and 30%.

Although physical buttons are the mainstream preference, the investigation into operational experience reveals potential cognitive load issues. Data shows that a significant proportion of users, when operating physical buttons (43.3%) and display screens (51.2%), indicate a need to "rely on the manual" to successfully complete operations (including Agree and Neutral options). Furthermore, the survey on design optimization needs points out that over 50% of users agree that "Simplified menus" (58.5%), "Clearer icons" (59.5%), and "One-touch shortcut keys" (55.4%) can effectively improve operational efficiency. This reflects that while existing interfaces can satisfy basic playing, there is still significant room for improvement in the intuitiveness of function navigation and information feedback.

3.2.3 Cluster Analysis

To deeply explore the heterogeneous needs of different user groups, this study employed Two-step Cluster Analysis. Based on variables such as learning history, usage frequency, self-assessed skill, and breadth of function usage, the 390 participants were divided into three groups with statistically significant differences. The three-cluster solution was determined by the Two-Step Cluster auto-clustering procedure, which showed the largest ratio of distance measures (2.72)—substantially higher than any other candidate solution. Analysis of Variance (ANOVA) results show that these three clusters exhibit significant differences in learning experience ($F=14.25$, $p<.001$), usage frequency ($F=16.76$, $p<.001$), skill level ($F=12.70$, $p<.001$), and function usage frequency ($F=42.81$, $p<.001$). The only variable with no significant difference was "Satisfaction," indicating that while different groups have different usage behaviors, subjective satisfaction with existing products tends to be consistent. (Table 1)

Table 1. One-way ANOVA and Post Hoc Results for Cluster Differences

Variable	Cluster 1 (n=186)	Cluster 2 (n=91)	Cluster 3 (n=113)	F(2,387)	η^2	Significant Post-hoc Differences (Scheffé)
Q7. Learning experience, M (SD)	2.39 (1.42)	1.75 (1.24)	1.64 (1.13)	14.25***	.069	1 > 2, 3
Q8. Usage frequency, M (SD)	2.45 (1.28)	1.97 (1.20)	1.67 (1.02)	16.76***	.075	1 > 2, 3
Q9. Self-assessed skill	1.80 (0.95)	1.46 (0.78)	1.33 (0.67)	12.70***	.059	1 > 2, 3
Q16. Feature usage frequency, M (SD)	2.96 (0.70)	2.35 (0.99)	2.11 (0.90)	42.81***	.175	1 > 2, 3
Q17. Satisfaction, M (SD)	3.59 (0.50)	3.54 (0.58)	3.53 (0.54)	0.64	.003	n.s.
Q18. Ease of use, M (SD)	3.62 (0.50)	3.47 (0.59)	3.33 (0.61)	9.26***	.046	1 > 3

Note. Values are presented as Mean (Standard Deviation). 1 = Cluster 1 (Advanced Expert Users); 2 = Cluster 2 (Intermediate Users); 3 = Cluster 3 (Novice Entry Users). η^2 = effect size. n.s. = not significant. *** = $p < .001$

3.2.4 User Profiles

The first group is "Advanced Expert Users" (n=186). This group accounts for the largest number of samples and is characterized by having the highest playing skills and usage frequency. They are extensively involved in diverse contexts such as teaching, creation, and ensemble performance. In terms of interface preference, they demonstrate the highest flexibility; not only are they familiar with physical buttons, but their usage rates for touch interfaces, screens, and Apps are also significantly higher than other groups, especially in music production and teaching contexts.

The second group is defined as "Intermediate Users" (n=91). Composed mainly of users with low-to-mid-level skills, their usage context is highly concentrated on "Personal Practice." The most significant characteristic of this group is their extreme reliance on "physical buttons"; in practice contexts, the proportion preferring physical buttons is as high as 96.7%, and they rarely use Apps or touch interfaces. This result suggests that for groups focused on practice, physical tactile feedback remains an indispensable core of interaction.

The third group is "Novice Entry Users" (n=113). This group consists mostly of beginners, and a proportion of 45.1% do not yet own a digital piano. They have the lowest function usage frequency and state "no involvement" in most advanced contexts (such as teaching, production).

Table 2 Cluster Labels and Descriptions

Cluster	Label	Key Characteristics	Interpretation
1 (n = 186)	Advanced Expert Users	• Highest levels of experience, frequency, and skill • Involved in the widest range of contexts (practice, teaching, production, ensemble) • Broad interface usage: physical buttons, touch, display, apps • Higher ownership of Yamaha and KAWAI, low proportion without a piano.	Experienced users with multi-purpose engagement, familiar with various digital piano functions, and versatile in teaching, practice, and creative production.
2 (n = 91)	Intermediate Users	• Lower skill/experience • Primarily focused on practice • Very strong reliance on physical buttons across contexts (practice, teaching, production) (80–97%) • Minimal use of touch, screen, or app interfaces.	Functional and efficiency-driven group preferring traditional physical controls; limited interest in advanced or digitalized features.
3 (n = 113)	Novice Entry Users	• Lowest learning experience and skill • Very limited engagement in teaching, production, ensemble, or other genres • Limited to basic practice contexts • High proportion of users with no owned piano (45.1%). • Rare use of advanced features, apps, or creative functions	Beginners or casual players, often without instrument ownership; lower engagement with digital features, focusing mainly on basic learning or occasional use.

4 DISCUSSION

Synthesizing the practical insights from the Phase One expert interviews and the quantitative data from the Phase Two large-scale questionnaire, this study follows the ISO 9241-210 Human-Centered Design process and incorporates considerations of user psychological expectations from Kansei Engineering to summarize a three-layer design framework comprising the "Functional Layer," "Interaction Layer," and "Emotional Layer." The following discussion delves into the design implications and empirical basis for each layer.

The Functional Layer is located at the bottom of the framework, covering hardware specifications, core technical capabilities, and information architecture. This study found a significant phenomenon of "function idleness" in existing digital pianos. Even for functions considered standard equipment, such as Transpose and Recording, over 50% of users stated they "rarely" or "never" use them. This data reveals an asymmetry between function supply and actual demand: when high-frequency functions (such as the metronome) are buried in secondary menus while low-frequency functions occupy the physical panel, the Interaction Cost exceeds the usage benefit, leading to excessive "Extraneous Cognitive Load" (Sweller, 2011) in the interface design.

Furthermore, cluster analysis results confirmed that users are not a homogeneous group. Although flagship models like the Yamaha P-525 attempt to satisfy all groups with a single interface, this "one-size-fits-all" strategy creates the design conflict described by Norman (2013): complex panel configurations deter Novice Entry Users (Cluster 3), while fixed layouts limit the operational efficiency of Advanced Expert Users (Cluster 1). Therefore, this study advocates that the Functional Layer should adopt "Frequency-Based Priority" and "Software Externalization" strategies. By transferring complex editing functions to Apps and allowing the physical interface to be dynamically defined based on user frequency, this approach can not only reduce the frustration caused by "Menu Diving" but also effectively reconcile the conflicting demands regarding operation depth between different clusters (Cluster 1 vs. Cluster 3).

The Interaction Layer focuses on the fluidity of the human-machine interface and the software-hardware integration experience. Research data shows a strong "Tactile Dependency," particularly in the "Personal Practice" scenario, where as many as 71% of users prefer physical buttons; notably, "Intermediate Users (Cluster 2)" exhibit a preference for physical interfaces as high as 96.7%. This supports the "Embodied Cognition" perspective of Schiavio and Menin (2013), which posits that a performer's thinking is closely linked to muscle memory, and purely touch-based interfaces lacking Haptic Feedback can easily interrupt the flow of performance.

However, this does not mean digital interfaces should be completely abandoned. Interview results reveal that users still rely heavily on visual confirmation during operation (e.g., confirming whether a function is active). Therefore, ideal interaction design should move towards a "Mixed Reality" mode, adopting a "Sync Feedback" strategy: retaining the physical damping of physical knobs to maintain performance intuition while providing real-time, large dynamic numerical feedback through digital screens. This design strategy resolves the physical coupling problem proposed by Hunt and Kirk (2000), satisfying modern users' dual needs for information transparency and digital convenience while preserving the feel of traditional instruments.

The Emotional Layer involves the user's psychological feelings and value judgments and is key to determining product satisfaction. The study found that users' pursuit of "authenticity" stems from a longing for traditional pianos, while the "headphone function" solves the psychological burden of fearing to disturb neighbors in modern living environments. From the perspective of Kansei Engineering, when visual expectations match actual tactile feedback, it can significantly enhance the emotional experience (Yanagisawa & Takatsuji, 2015).

Addressing the inherent limitation of digital instruments lacking a physical hammer structure, this study suggests using "Visual Compensation" to enhance Perceived Value. For example, presenting simulated physical dynamics (such as virtual hammer strike animations) through software interfaces can use visual information to compensate for the lack of physical entities, establishing a deeper emotional connection. The design implication of this layer is that digital pianos should not be seen merely as a stack of functions, but as carriers satisfying user psychological needs and identity, thereby elevating overall brand loyalty.

Taken together, these three layers translate into concrete design guidelines. At the Functional Layer, high-frequency functions identified in this study—such as the metronome and tone adjustment—should remain as dedicated physical controls, while low-frequency functions such as Transpose and Recording can be externalized to a companion app, reducing panel clutter without sacrificing accessibility. At the Interaction Layer, a Sync Feedback design—for example, a physical volume knob paired with a real-time numerical readout on an adjacent screen—preserves tactile intuition while resolving the visual confirmation needs identified in the interviews. At the Emotional Layer, Visual Compensation techniques, such as simulated hammer-strike animations, can enhance perceived authenticity despite the absence of a physical hammer mechanism. Together, these guidelines address the divergent needs of Advanced, Intermediate, and Novice user clusters identified in this study.

5 CONCLUSION

This study reveals the behavioral differences among distinct user clusters (Advanced, Intermediate, Novice) and proposes a three-layer Design Framework integrating "Functional, Interaction, and Emotional" aspects for digital piano interfaces. Future research is suggested to develop prototypes with reconfigurable interface functions to conduct usability testing, aiming to measure the specific effectiveness of this framework in reducing cognitive load and improving operational performance in actual operational tasks.

REFERENCES

Hunt, A., & Kirk, R. (2000). Mapping Strategies for Musical Performance. *Trends in Gestural Control of Music*, 231–258.

Jensenius, A. R. (2022). *Sound Actions: Conceptualizing Musical Instruments*. The MIT Press.

Jorritsma, W., Cnossen, F., & Van Ooijen, P. M. A. (2015). Adaptive support for user interface customization: A study in radiology. *International Journal of Human-Computer Studies*, *77*, 1–9. https://doi.org/10.1016/j.ijhcs.2014.12.008

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

McPherson, A. (2012). TouchKeys: Capacitive Multi-touch Sensing on a Physical Keyboard. In A. R. Jensenius & M. J. Lyons (Eds.), *A NIME Reader* (Vol. 3). Springer International Publishing. https://doi.org/10.1007/978-3-319-47214-0_27

Norman, D. A. (2013). *The design of everyday things* (Rev. and expanded edition). Basic Books.

Schiavio, A., & Menin, D. (2013). *Embodied Music Cognition and Mediation Technology*: A critical review. *Psychology of Music*, *41*(6), 804–814. https://doi.org/10.1177/0305735613497169

Shaer, O., & Hornecker, E. (2010). Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends® in Human–Computer Interaction*, *3*(1–2), 1–137. https://doi.org/10.1561/1100000026

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Welter, L., & Ren, S. (2025). *Yamaha P-525 review: In-Depth Look At The New Flagship*. https://www.pianodreamers.com/yamaha-p525-review/

Wessel, D., & Wright, M. (2002). *Problems and Prospects for Intimate Musical Control of Computers*.

Yamaha. (2023). *P-525—Overview—P Series—Products—Pianos—Musical Instruments—Yamaha—Other European Countries & Regions*. https://europe.yamaha.com/en/musical-instruments/pianos/products/p-series/p-525/

Yanagisawa, H., & Takatsuji, K. (2015). Effects of Visual Expectation on Perceived Tactile Perception: An Evaluation Method of Surface Texture with Expectation Effect. *International Journal of Design*, *9*(1), 39–51.

RELAXATION EFFECTS OF VISUAL FEEDBACK FROM A BREATH-SYNCHRONIZED INTERACTIVE DEVICE: A COMPARISON OF SYNCHRONIZED AND DELAYED CONDITIONS

Rina Katsube[a], Toshiaki Uchiyama[a], Shinichi Koyama[a*]

[a] *University of Tsukuba, Tsukuba, Japan*

ABSTRACT

Sensory sensitivities are prevalent across a wide range of individuals and can cause stress and fatigue (Horder et al., 2014). Sensory rooms—spaces designed to regulate sensory input and promote relaxation—employ items such as bubble tubes (Su et al., 2025). Recently we developed a device equipped with an RGB-Depth camera to detect respiration in real time that deforms a hyperboloid structure in response to breathing (Katsube et al., 2025). However, the mechanisms underlying their relaxation effects remain poorly understood (Su et al., 2025). This study tested two hypotheses. First, based on predictive coding theory (Friston, 2010), we hypothesized that the perception of regular, predictable motion promotes relaxation. Second, we hypothesized that visual feedback synchronized with one's own breathing further enhances relaxation (Betka et al., 2023; Blum et al., 2020). To test these hypotheses, we tested our interactive device in synchronized and delayed feedback conditions. In the experiment, 54 participants (aged 18–35) were randomly assigned to a synchronized (real-time feedback) or delayed (2.5-second lag) group. We measured respiratory rate and amplitude, heart rate variability (HRV), salivary amylase, and self-report measures; sensory profiles were assessed using the Adolescent/Adult Sensory Profile (AASP) to confirm group equivalence. Both groups exhibited relaxation effects; however, the synchronized group demonstrated significantly greater reductions in respiratory rate (from 15.6 to 10.9 breaths/min, sustained throughout the session), approximately 1.7-fold increases in breathing amplitude, and decreased salivary amylase levels compared to the delayed group. lnRMSSD increased over time in both groups, with no significant group difference. That both groups relaxed is consistent with the first hypothesis that predictable, regular motion promotes relaxation (Friston, 2010). The stronger effects in the synchronized group are consistent with the second hypothesis that breathing synchronization further enhances relaxation (Betka et al., 2023; Blum et al., 2020). The results offer design implications for personalized sensory environments.

Keywords: *breath-synchronized feedback, predictive coding, relaxation, sensory room, respiratory biofeedback*

[*] Corresponding author: Shinichi Koyama, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan; skoyama@geijutsu.tsukuba.ac.jp

1 INTRODUCTION

Sensory sensitivities are prevalent across a wide range of individuals and can cause stress and fatigue (Horder et al., 2014). Sensory rooms employ items such as bubble tubes to promote relaxation, but the mechanisms by which they induce relaxation remain poorly understood (Su et al., 2025). One factor is stimulus predictability. According to predictive coding theory (Friston, 2010), the brain minimizes prediction error—the discrepancy between predicted and actual sensory input. Stimuli with regular, predictable movements may reduce prediction error and promote relaxation. Another factor is breathing regulation: breathing patterns are closely related to emotional states, and external stimuli synchronized with breathing could enhance breathing awareness.

We developed a breath-synchronized interactive device that detects respiration via an RGB-Depth camera and deforms a hyperboloid structure accordingly (Katsube et al., 2025). This study compared synchronized (real-time) and delayed (2.5-s lag) feedback to test two hypotheses: (H1) regular, predictable motion promotes relaxation regardless of synchronization, and (H2) breath-synchronized feedback further enhances relaxation.



Figure 1. The breath-synchronized interactive device. (a) System components: hyperboloid structure, Kinect v2, and motor. (b) Deformation during inhalation and exhalation.

2 METHODS

54 participants (38 women, 16 men; aged 18–35; M = 24.6) were randomly assigned to a synchronized (n = 27) or delayed group (n = 27) based on an a priori power analysis (f = 0.25, α = .05, 1 – β = .95). The device (Figure 1) comprised a Kinect v2 camera, a motor, and a hyperboloid structure that twisted during inhalation and reversed during exhalation. In the synchronized group, the device responded in real time; in the delayed group, with a 2.5-s lag. After a 5-minute rest, participants viewed the device for 15 minutes while breathing naturally. Respiratory rate, amplitude, and heart rate variability (lnRMSSD, lnHF; Polar H10) were measured at four time points (rest 2–5 min; viewing 2–5, 7–10, 12–15 min); salivary amylase was measured before and after viewing. Mixed ANOVA (time × group) was used for physiological indices; nonparametric tests for salivary amylase. After viewing, participants completed self-report ratings including overall comfort; sensory profiles were assessed using the AASP (Brown & Dunn, 2002) with no significant group difference (Pillai's Trace = 0.032, p = .802). Exploratory associations between sensory characteristics and subjective responses were examined using Spearman's rank correlations. The study was approved by the institutional ethics committee (GEI025-1).

3 RESULTS

3.1 Respiratory rate and amplitude

The time × group interaction was significant for both respiratory rate ($F(2.47, 128.64) = 4.95$, $p = .005$, $\eta p^2 = .087$) and amplitude ($F(3, 156) = 3.71$, $p = .013$, $\eta p^2 = .067$). The synchronized group's respiratory rate decreased from 15.6 to 10.9 breaths/min (2–5 min), remaining at 11.5 (12–15 min), whereas the delayed group showed only an initial decrease (14.9 → 12.3 → 13.2 breaths/min). Amplitude increased approximately 1.7-fold in the synchronized group (5.1 → 8.5–8.9) versus a smaller change in the delayed group (4.2 → 5.4–5.6; Figure 2).

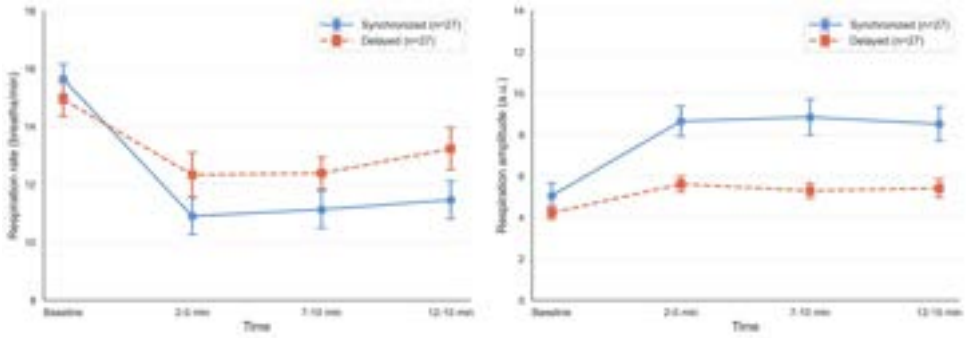


Figure 2. Respiratory rate (left) and amplitude (right) across time points. Error bars indicate standard errors.

3.2 HRV

For lnRMSSD, the main effect of time was significant ($p = .033$), but the time × group interaction ($p = .197$) and group effect ($p = .899$) were not. For lnHF, no significant effects were found.

3.3 Salivary amylase

Salivary amylase decreased significantly in the synchronized group (from 19.0 to 14.0 kIU/L; $W = 220.0$, $p = .013$, $r = .594$) but not in the delayed group (Figure 3). The between-group difference was significant ($U = 533.5$, $p = .003$, $r = -.464$).

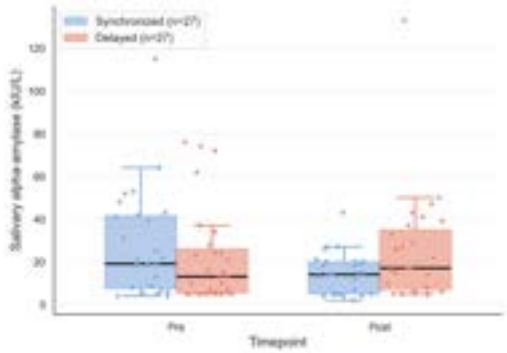


Figure 3. Salivary amylase levels (kIU/L) before and after viewing. Thick lines indicate medians.

3.4 Associations with sensory characteristics

In exploratory analyses, higher Sensory Sensitivity scores on the AASP were associated with lower overall comfort in the delayed group ($rs = -.396$, $p = .041$) but not in the synchronized group, while higher Low Registration scores were associated with greater comfort in the delayed group ($rs = .387$, $p = .046$).

4 DISCUSSION AND CONCLUSION

Both conditions induced relaxation-related changes, suggesting that the device's regular motion may have contributed to reduced prediction error (Friston, 2010), consistent with H1. The stronger effects in the synchronized group can be interpreted through Seth's (2013) interoceptive predictive coding model and are consistent with previous findings (Betka et al., 2023; Blum et al., 2020), supporting H2. These findings suggest that synchronized feedback may be suitable as a default mode for sensory environments, as delayed feedback may reduce comfort among users with higher sensory sensitivity. Future studies should examine clinical populations, broader age ranges, longer-term use, and larger samples for individual-difference analyses.

ACKNOWLEDGMENTS

This study was supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (C) to Rina Katsube, Toshiaki Uchiyama, and Shinichi Koyama (Grant Number: 25K15774).

REFERENCES

Betka, S., Kannape, O. A., Fasola, J., Lance, F., Cardin, S., Schmit, A., Similowski, T., Soccal, P. M., Herbelin, B., Adler, D., & Blanke, O. (2023). Virtual reality intervention alleviates dyspnoea in patients recovering from COVID-19 pneumonia. ERJ Open Research, 9(6), Article 00570-2022. https://doi.org/10.1183/23120541.00570-2022

Blum, J., Rockstroh, C., & Goritz, A. S. (2020). Development and pilot test of a virtual reality respiratory biofeedback approach. Applied Psychophysiology and Biofeedback, 45(3), 153–163. https://doi.org/10.1007/s10484-020-09468-x

Brown, C., & Dunn, W. (2002). Adolescent/adult sensory profile: User's manual. Pearson Education Limited.

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Horder, J., Wilson, C. E., Mendez, M. A., & Murphy, D. G. (2014). Autistic traits and abnormal sensory experiences in adults. Journal of Autism and Developmental Disorders, 44(6), 1461–1469. https://doi.org/10.1007/s10803-013-2012-7

Katsube, R., Uchiyama, T., & Koyama, S. (2025). Development of a breath-synchronized interactive device for sensory rooms. In C.-Y. Chang, C.-H. Chen, & Y. Hsu (Eds.), IASDR 2025: Design Next. Design Research Society. https://doi.org/10.21606/iasdr.2025.925

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Su, A., Shen, J., & Koyama, S. (2025). Exploring the relationship between preferred bubble tube speeds in sensory rooms and physiological–psychological factors. F1000Research, 14, 214. https://doi.org/10.12688/f1000research.161090.1

SIMPLIFICATION OF USER PREFERENCE RULES FOR PERSONALIZED MAKEUP RECOMMENDATION IN AN AUTOMATIC VIRTUAL MAKEUP SYSTEM

Yui Tokushima[a], Hiroshi Takenouchi[b]

[a] *Graduate School of Engineering, Fukuoka Institute of Technology, Japan, mjm26104@bene.fit.ac.jp*

[b] *Fukuoka Institute of Technology, Japan, h-takenouchi@fit.ac.jp*

ABSTRACT

This study introduces a method for simplifying user preference rules in a virtual makeup system while maintaining user satisfaction with the generated results. Preference rules extracted using interactive evolutionary computation and fuzzy inference often contain redundant antecedent conditions, which reduce interpretability. To address this issue, we propose a rule-simplification method based on cosmetic feature importance, defined by their frequency of appearance across generations. The performance of the proposed approach is evaluated in experiments involving 16 participants, comparing user satisfaction with preference rules generated from rule sets containing 3, 5, 7, and 10 conditions. The results suggest that users' makeup preferences can be represented with fewer conditions without a significant decrease in satisfaction. These findings indicate that the proposed method preserves the essential characteristics of user preferences and is particularly effective in simplifying complex "dislike" preference rules.

Keywords: Interactive evolutionary computation, winner-based paired comparison method, genetic algorithm, preference rule simplification, virtual makeup.

1 INTRODUCTION

Recent years have witnessed increased use of online shopping and online counseling across a wide range of age groups, attributable to the widespread adoption of smartphones and impact of the COVID-19 pandemic. This has accelerated digital transformation in the cosmetic domain, including the deployment of virtual makeup systems that bridge the gap between digital representations and real products (Yupelmi et al., 2023). Such systems have been shown to reduce consumer anxiety during product trials and to enhance user engagement and brand trust (Micheletto et al., 2025).

Despite these advantages, most existing virtual makeup systems require users to manually select and combine individual cosmetic items, which are time-intensive and require considerable cognitive effort. To overcome this limitation, automatic makeup generation methods that

consider facial features and personal attributes have been proposed (Nguyen & Liu, 2017). However, these approaches provide limited support for explicitly representing and explaining user preferences in a human-interpretable form.

Interactive evolutionary computation (IEC) has been applied to virtual makeup systems to interactively acquire user preferences (Takenouchi et al., 2023). To mitigate user fatigue caused by repeated evaluations, paired-comparison methods have been introduced in IEC and shown to be effective for preference extraction (Takagi & Pallez, 2009). Winner-based paired comparison (WPC) represents a practical implementation of paired comparison in IEC-based virtual makeup systems. Building on this framework, Takenouchi et al. (2025) proposed a preference-rule-generation mechanism based on fuzzy inference. Preference rules make it possible to identify the cosmetic features associated with each user's evaluations and provide reasons for personalized recommendations. However, redundant antecedent conditions make it difficult to identify which features primarily affect each recommendation.

To address these limitations, this paper introduces a method for simplifying preference rules generated in an IEC-based makeup support system using the WPC method. Shorter rules enable users and system designers to identify influential cosmetic features more easily and understand recommendation results with fewer conditions. Unlike methods that explain prediction results after model construction, the proposed method directly simplifies preference rules obtained through IEC based on antecedent occurrence frequency. Its contribution is an interpretable Kansei-based framework that reduces redundant conditions while maintaining subjective satisfaction. The proposed method was evaluated through a user study with 16 participants to examine whether reducing the number of antecedent conditions affects user satisfaction.

2 PROPOSED SYSTEM

2.1 System overview

Figure 1 schematically illustrates the proposed IEC-based virtual makeup system using the WPC method. Makeup patterns evolve through user evaluations using a genetic algorithm (GA), and "like" and "dislike" preference rules are extracted from individuals with high and low fitness values. For example, a simplified preference rule with three antecedent conditions can be expressed as follows: if the foundation is dark, the lipstick is dark, and the eyeshadow is light, then the user evaluates the makeup as "like."

This study focuses on the rule-simplification module, which improves interpretability by reducing redundant antecedent conditions based on feature frequency. Minor refinements to makeup transparency thresholds are introduced to enhance visual realism.

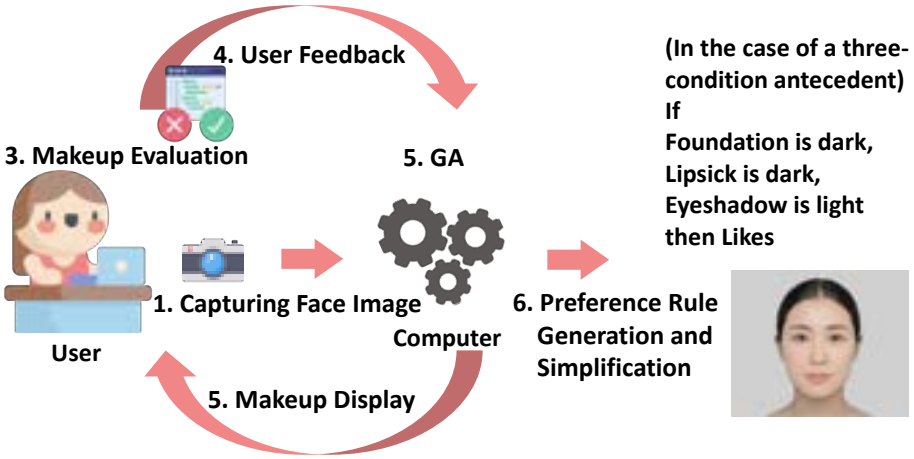


Figure 1. Overview of a Virtual Makeup System Incorporating Preference Rule Generation and Simplification Using IEC and WPC

2.2 Gene coding of makeup parts

Figure 2 illustrates the makeup configuration and its encoding scheme. The colors and opacity levels of six makeup parts—foundation, lips (lipstick), eyeshadow, cheeks (blush), eyebrows, and eyeliner—are encoded in a 28-bit gene string, resulting in a large combinatorial search space. Figure 3 illustrates the individual makeup parts used in the proposed system.

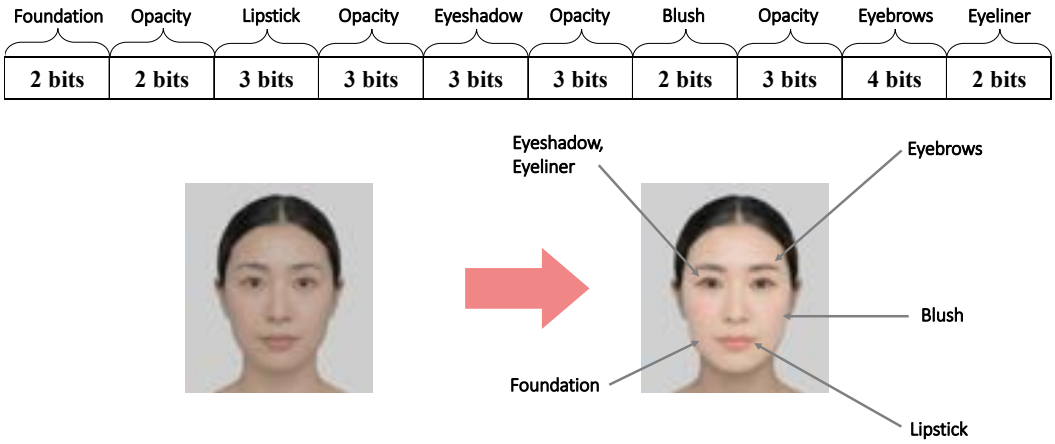


Figure 2. Makeup Configuration

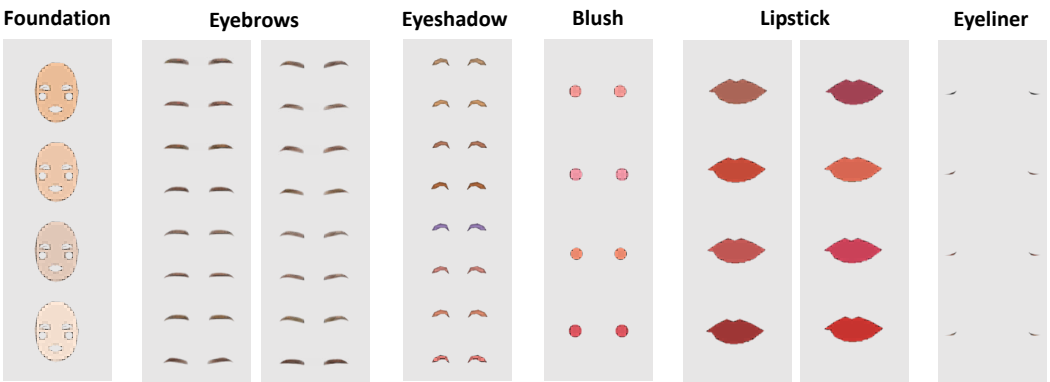


Figure 3. Makeup parts

2.3 Individual evaluation using WPC

Figure 4 illustrates the evaluation protocol within the WPC method. In WPC, individuals are evaluated through paired comparisons: two individuals are presented to the user, and evaluation scores are iteratively updated based on the comparison results to obtain a relative ranking.

All individuals are initially assigned an evaluation score of 1. In each comparison, the user selects the preferred individual, and the winner's score is updated according to predefined rules, while the other individual is replaced. This process is repeated until all individuals have been evaluated.

Table 1 summarizes the score-update rules, including cases where the user selects "both like" or "both dislike." These rules adjust evaluation scores according to relative preference consistency, following previous studies (Takenouchi et al., 2025).

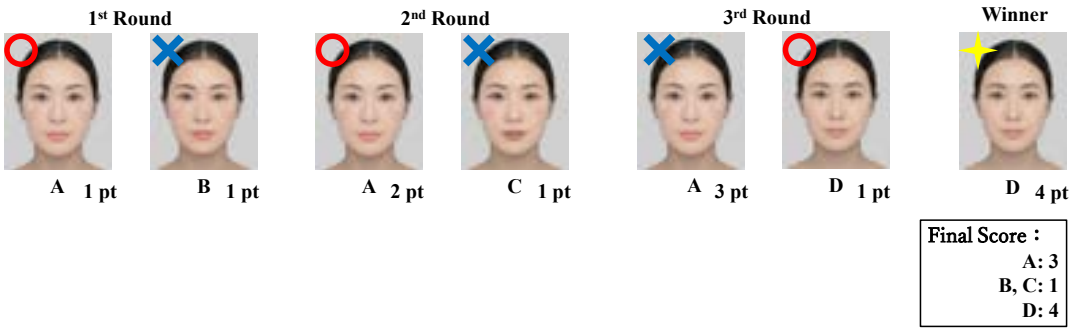


Figure 4. Individual evaluation method

Table 1. Changes in evaluation scores when "both like" or "both dislike" is selected

		Both like	**Both dislike**
Evaluation scores of presented makeup patterns	Match	Scores of both patterns +1	Scores of both patterns -1
	Mismatch	Lower score →Higher score	Higher score →Lower score

2.4 Fuzzy rule structure

The fuzzy rules used in this study adopt a general structure consisting of antecedent and consequent parts. The antecedent assigns binary labels to each feature, and the consequent takes binary evaluation labels. The number of antecedent features is set as $M = 10$. The rules are defined as

$$R_i = (R_i^1, R_i^2, \ldots, R_i^M, O_i) \quad (1)$$

where R_i^j denotes the antecedent label of the j-th feature in the i-th fuzzy rule, and O_i represents the consequent label of the corresponding rule.

2.5 Membership function design

Following previous studies (Takenouchi et al., 2025), two-label membership functions are defined for all makeup parts and their opacity, with minor modifications to reflect makeup design

characteristics. For foundation, eyebrow, and eyeliner, dark/bright labels are used, whereas subtle/vivid labels are assigned to lipstick, eyeshadow, and blush. The opacity of selected makeup parts is represented by light and heavy labels. Figure 5 shows the lipstick membership function as a representative example. The consequent membership functions consist of “dislike” and “like” labels.

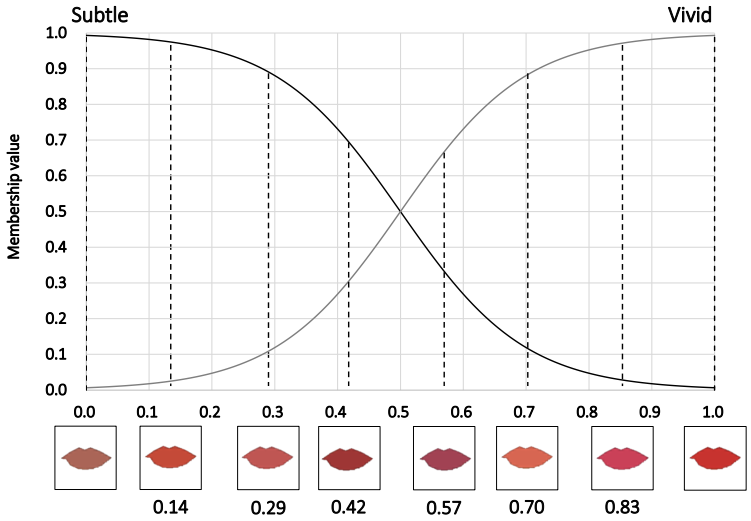


Figure 5. Lipstick membership function

2.6 Preference rule generation and organization

User preferences are represented as a set of fuzzy rules based on membership functions defined for the color and opacity of each makeup part. Each preference rule consists of ten antecedent items: six related to makeup colors and four related to opacity levels, with membership values assigned based on user evaluations.

To investigate the minimum number of antecedent conditions required to adequately represent user satisfaction, multiple rule sets with different numbers of antecedent conditions are constructed and evaluated. This allows us to examine whether reducing the number of conditions affects user satisfaction with the generated preference rules.

In each generation of the IEC process, “like” and “dislike” rules are generated from highly rated and poorly rated makeup patterns, respectively, thereby accumulating preference rules over successive generations.

To reduce redundancy, antecedents are first compared among rules of the same type, i.e., among “like” and “dislike” rules. When three or more of five selected antecedent items (excluding foundation and opacity) share identical membership values, the rules are considered identical and the older rule is compared.

Foundation-related parameters are excluded from this comparison because foundation color tends to exhibit less variation in user preference compared with other makeup parts, and its influence on distinguishing individual preferences is relatively limited.

Next, antecedents are compared between “like” and “dislike” rules. When all six color-related antecedent items match, the rules are considered contradictory and both are removed to maintain consistency. In this manner, we obtain a compact and consistent set of preference rules that reflect user preferences while eliminating redundant and contradictory information.

2.7 Preference-rule-simplification method

This section outlines the proposed rule-simplification method based on reducing the number of antecedent conditions. The importance of each antecedent condition is defined by its frequency of occurrence in the generated preference rules. Conditions with higher importance are retained as the top K antecedents, where $K = 3, 5, 7, or\ 10.$

Let R be the set of all preference rules and c an antecedent condition. The frequency $f(c)$ is defined as

$$f(c) = \sum_{r \in R} I(c \in r) \tag{2}$$

where $I(c \in r)$ is an indicator function that takes the value 1 if the antecedent label c appears in rule r, and 0 otherwise. Thus, $f(c)$ represents the frequency of occurrence of antecedent label c across all extracted preference rules. The importance weight $w(c)$ is obtained by normalizing $f(c)$:

$$w(c) = \frac{f(c) - \min f(c)}{\max f(c) - \min f(c)} \tag{3}$$

Each preference rule originally contains ten antecedents, from which only the top K conditions ranked by $w(c)$ are retained.

2.8 Evaluation interface

Figure 6 shows the evaluation interface used in this study. The user examines two makeup candidates and adjusts the opacity of each makeup part using sliders. Subsequently, the user selects the preferred makeup or chooses both candidates as “like” or both as “dislike.” Based on this input, the system calculates evaluation scores and automatically presents the next pair of makeup candidates. The user repeats this evaluation process for one generation. After the evaluations for a generation are completed, the system applies a GA to generate the next population of makeup candidates. This process is repeated until the final generation is reached. After completing the evaluations of the final generation, the system simplifies the extracted preference rules and generates the final set of simplified preference rules. Next, the system sequentially presents the preference rules with different numbers of antecedent conditions (3, 5, 7, and 10) to the same user, who evaluates each rule using a five-point rating scale.



Figure 6. Evaluation interface after facial image capture

3 EVALUATION EXPERIMENT

3.1 Overview

A user study was conducted to evaluate the proposed preference-rule-simplification method using a GA-based makeup design system. This study was approved by the ethics committee of Fukuoka Institute of Technology, and informed consent was obtained from all participants. Preference rules were generated from participants' evaluations and simplified by reducing the number of antecedent conditions. Table 2 summarizes the experimental settings. Sixteen female university students evaluated simplified preference rules containing 3, 5, 7, or 10 antecedents on a five-point scale. This exploratory study used a small sample because each participant repeatedly evaluated makeup patterns over ten IEC generations; therefore, the findings should be regarded as preliminary. The results were analyzed to assess the effect of antecedent reduction on user satisfaction.

Table 2. Experimental settings

Item	Description
Participants	16 female university students
Experimental environment	Same display used for all participants
Population size	10 individuals per generation
Number of generations	10
Gene length	28 bits
Mutation rate	15%
Crossover method	Uniform crossover
Selection method	Roulette selection with elitism
Antecedents (K)	3, 5, 7, 10

3.2 Results of rule generation

Figure 7 shows the number of generated preference rules across all participants. On average, the number of "dislike" rules (9.75) was larger than that of "like" rules (4.50). This suggests that participants have more diverse or clearer criteria for unattractive makeup patterns.

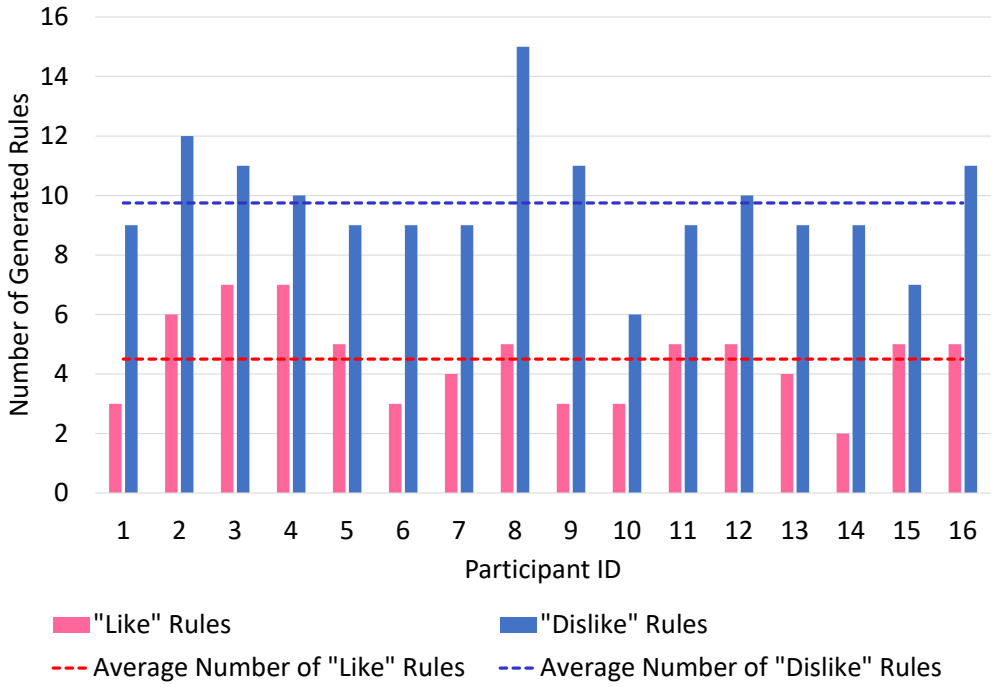


Figure 7. Number of generated preference rules

3.3 Satisfaction levels

Figure 8 illustrates the satisfaction levels for the simplified rules. For “like” and “dislike” rules, the mean scores were 3.08–3.48 and 4.04–4.08, respectively. The Friedman test revealed no significant difference in satisfaction among the four rule sets for either the “like” or “dislike” rules.

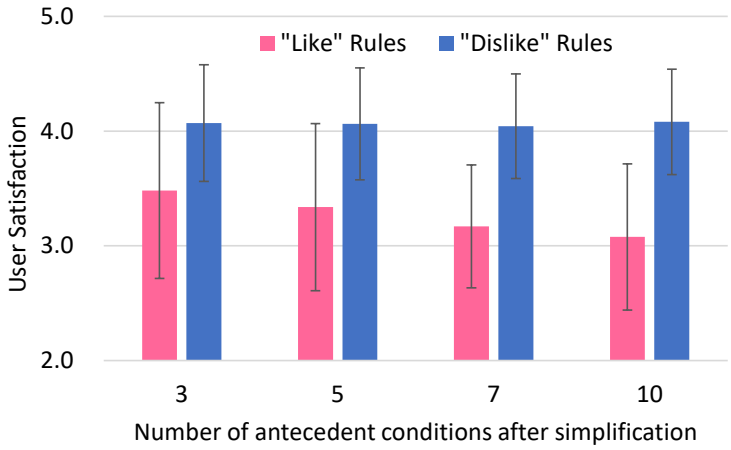


Figure 8. Mean and standard deviation of satisfaction for each preference rule across all participants

4 DISCUSSION

The experimental results showed that the number of generated “dislike” rules was consistently higher than that of “like” rules across all participants, indicating that users have clearer and more diverse criteria for identifying undesirable makeup than for desirable makeup. This suggests that

negative preferences may produce more redundant and complex rule sets than positive preferences.

In terms of user satisfaction, no significant differences were observed among rule sets with different numbers of antecedent conditions for either “like” or “dislike” rules. This indicates that reducing the number of antecedent conditions does not degrade user satisfaction. Moreover, “dislike” rules consistently received higher satisfaction scores than “like” rules, suggesting that simplified rules are more easily accepted when users’ evaluation criteria are explicit.

Simplified rules could provide concise recommendation reasons in virtual makeup, AR, beauty retail, and online consultation services. They may support understanding, personalization, and trust, although these user-centered effects were not directly evaluated.

Overall, these results demonstrate that the proposed simplification method improves the interpretability of preference rules while preserving user preferences, making it suitable for addressing redundancy in conventional preference-rule-generation methods. Thus, this study offers a novel empirical framework that ensures the interpretability of preference rules derived from IEC while retaining only a minimal set of essential information, even when the original rules become overly complex.

5 CONCLUSION

This study established a preference-rule-simplification method for an IEC-based virtual makeup system to improve interpretability. The importance of antecedent conditions was defined using their frequency of occurrence in the preference rules generated during the IEC process, and only high-importance conditions were retained. The results highlighted that users’ preference judgments can be effectively represented using a limited set of cosmetic features. The proposed method was particularly effective in simplifying redundant “dislike” preference rules. From a Kansei Engineering perspective, the method concisely represents subjective preferences using influential cosmetic attributes. However, the occurrence frequency of antecedent conditions does not capture interactions among cosmetic features, and the controlled study with 16 students limits generalizability. Future work will involve larger user studies, comparisons with alternative rule-simplification and explanation methods, including explainable AI approaches, importance measures that consider interactions among cosmetic features, and real-world cosmetic applications.

REFERENCES

Micheletto, V., Accardi, S., Fici, A., Piccoli, F., Rossi, C., Bilucaglia, M., Russo, V., & Zito, M. (2025). Enjoy it! Cosmetic try-on apps and augmented reality: The impact of enjoyment, informativeness and ease of use. Frontiers in Virtual Reality, 6–1515937. https://doi.org/10.3389/frvir.2025.1515937

Nguyen, T. V., & Liu, L. (2017). Smart mirror: Intelligent makeup recommendation and synthesis. Proceedings of the 25th ACM International Conference on Multimedia (MM ’17), 1609–1611.

Takagi, H., & Pallez, D. (2009). Paired comparison-based interactive differential evolution. Proceedings of the World Congress on Nature & Biologically Inspired Computing (*nabic)*, 475–480. https://doi.org/10.1109/NABIC.2009.5393359

Takenouchi, H., Isayama, S., & Tokumaru, M. (2023). Automatic Virtual Makeup System Using User-Preference Information. HCI International 2023 Posters, *1836*, 523–529. https://doi.org/10.1007/978-3-031-36004-6_71

Takenouchi, H., Nakajima, M., & Tokumaru, M. (2025). User-Preference Rule Generation in Automatic Virtual Makeup System. *HCI International 2025 Posters*, *2528*, 377–387. https://doi.org/10.1007/978-3-031-94168-9_37

Yupelmi, M., Yulastri, A., Ganefri, Giatman, H., Effendi, H., Muskhir, M., & Mukhlidi. (2023). A Comprehensive Exploration of Entrepreneurial Strategies in The Makeup and Beauty Industry: The Role of Social Media Marketing. *Indonesian Journal of Computer Science*, *12*(5), 2498–2513. https://doi.org/10.33022/ijcs.v12i5.3439

SIMULATING PERCEPTION: EXPLORING AN AI-ASSISTED APPROACH TO KANSEI ENGINEERING IN EMOTIONAL PRODUCT DESIGN

Sigrid ÅKERSKOG[a], Linus FROSTELL[a], Nuno VALVERDE[b], Simon SCHÜTTE[a], Lluís MARCO-ALMAGRO[1 c]

[a] *Linköping University, Sweden*
[b] *Logitech Europe SA, Switzerland*
[c] *Universitat Politecnica de Catalunya - BarcelonaTech, Spain*

ABSTRACT

Kansei Engineering translates emotional user responses into product design decisions, normally using self-reported questionnaires where users rate products on different perceptions (Kansei words). This paper investigates whether a large language model can support this data collection step by approximating users' emotional ratings of product images. A browser-based tool called KansAI has been developed, in different versions, to assess to what extent human ratings can be simulated by an engine powered by the Claude API. The study uses AI-generated computer mouse images as product stimuli and ten kansei words in the semantic space: Ergonomic, Sturdy, Sleek, Organic, Premium, Quick, Cheerful, Trustworthy, Grippy and Futuristic.

Two Kansei studies with humans were conducted: a first one for calibration of the system (with 71 respondents rating nine images of computer mice), and a second one for validation of the system (with 30 respondents rating five new images of computer mice). Four versions of the KansAI system were compared with progressively richer calibration information: no human data, demographic profiles, mean calibration ratings and mean ratings with few-shot individual examples. Both the aggregated and individual accuracy was evaluated using the mean absolute error between real human ratings and KansAI ratings.

Calibration with real human ratings improved aggregate prediction substantially. There were also differences depending on the type of Kansei word: KansAI performed best on Kansei Words with clearer sensory or visual cues, such as ergonomic, sturdy and sleek, and least accurately on symbolic or affective descriptors, such as futuristic, trustworthy and cheerful. However, the system was not able to correctly simulate the individual variability among human responses. The findings suggest that calibrated LLMs can complement human-centered Kansei Engineering as an early-stage exploratory support tool.

Keywords: *emotional product design; large language models; AI-assisted evaluation; computer mice*

[1] Corresponding author: lluis.marco@upc.edu

1 INTRODUCTION

Kansei Engineering (KE) is a method for translating users' subjective impressions and affective and emotional needs into design-relevant information. It is especially relevant when product differentiation depends not only on functional performance, but also on how a product is perceived (Nagamachi, 1995; Norman, 2003; Schütte et al., 2023). Traditional KE studies rely on developing the semantic space, with a selection of kansei words (KWs), the space of properties, with representative product samples, and linking both the semantic space and the space of properties using descriptive and statistical tools based on human respondent ratings. This human-centered foundation is essential, but it can make early-stage design evaluation slow and costly.

At the same time, multimodal large language models (LLMs) are becoming capable of interpreting both text and images. Recent work has explored LLMs as tools for building semantic spaces in Kansei studies, while broader research on AI agents suggests that calibrated models may approximate aspects of human survey behavior under specific conditions (Alcaide-Marzal & Diego-Mas, 2025; Park et al., 2025). These developments raise a practical question for affective design: can an AI tool be calibrated to approximate the aggregate emotional perception of a target group? And if so, can it also work at an individual level? And will it work equally well no matter the type of Kansei word used?

This paper presents the development of a tool, based on the Claude AI API, called KansAI, to simulate Kansei evaluations of product images. The study, developed in collaboration with the consumer electronics company Logitech, focuses on computer mice.

2 HUMAN KANSEI ENGINEERING STUDIES

This section describes how the traditional Kansei engineering study was prepared: selection of Kansei words, creation of products to be rated, and conduct of the data collection.

2.1 Selection of Kansei words

To gather Kansei words for the study, a hybrid approach combining manual work and AI was used. Relevant Kansei Words (KWs) for computer mice were first collected by analyzing review blogs, online discussions, and three academic papers on mouse design, with Gemini used to support text mining, synonym generation, and preliminary clustering. The expanded list was then manually reduced by removing semantically overlapping terms and reorganizing the remaining words into distinct semantic groups.

To validate this hybrid approach, a workshop was conducted with five LiU participants who interacted with five different computer mice and described the emotions and impressions each product evoked. The participants were instructed to avoid obvious physical descriptions, such as color, and the collected words were translated into English and semantically grouped for comparison with the generated KW list.

A final set of 10 Kansei words was selected for use in the studies: Ergonomic, Sturdy, Sleek, Organic, Premium, Quick, Cheerful, Trustworthy, Grippy and Futuristic.

2.2 Generation of product images

Images of computer mice were created using AI. An evaluation of state-of-the-art generative AI image tools was conducted to select the most suitable engine for product visualization in the project. The engines were evaluated according to photorealism and quality, prompt adherence, design coherence, aesthetic consistency, ease of use, and editing possibilities. These criteria are especially important for Kansei Engineering, where unrealistic or inconsistent visuals could influence participants' emotional perceptions. Gemini's Nano Banana 2 was selected as the primary image generation engine because of its strong overall performance and suitability for iterative prompt refinement. ChatGPT's GPT Image was used as a complementary tool during final refinement, as it showed stronger instruction-following when editing existing generated images. The two models were therefore used in parallel, leveraging their respective strengths.

To ensure a consistent baseline across all created images in terms of lighting, camera angle, and composition, an existing Logitech product photograph was used as a reference image. Following known methods (Barros, 2024), utilizing an input image in text-to-image models provides a structural foundation, allowing the designer to control the deviation from the original design by balancing the prompt against the reference image's geometry.

From the 14 different mouse images created, nine of them (shown in Figure 1) were used iteratively as calibration images in the AI study.



Figure 1. Mouse images (mice 1 to 9) used for calibration in the AI study

The other five images (shown in Figure 2) were set aside to be used for validation.



Figure 2. Mouse images (mice 10 to 14) used for validation

2.3 Data collection

Two questionnaires were used to collect user ratings for all Kansei words for each of the presented mouse images. The two studies serve two different purposes: the first one, with mice 1 to 9 (Figure 1), as calibration data for the AI tool; the second one, with mice 10 to 14 (Figure 2), as data for comparison with the results given by the AI – that is, as validation data.

The two surveys were built and distributed using KESo, a web-based tool developed specifically for conducting KE studies (Schütte et al., 2006). Besides presenting stimuli to be rated on all Kansei words in a randomized way, KESo allows the inclusion of demographic questions. Ratings ranged from 0 to 100, based on a visual-analogue scale.

The survey data were cleaned by removing demographic outliers, duplicate entries, and likely careless responses. Careless responses were identified through straight-lining, defined as leaving more than three consecutive ratings unchanged, and through extreme-response patterns where most Kansei ratings were at the scale extremes.

Table 1 summarizes the participants profiles for both studies.

Table 1. Calibration and validation Kansei engineering studies, done with KESo

Study	Stimuli used	Participants	Gender	Age
Calibration survey	9 (mice 1 to 9)	71	37 male, 34 female	31 aged 18–24; 40 aged 25–29
Validation survey	5 (mice 10 to 14)	30	15 male, 15 female	5 aged 18–24; 25 aged 25–29

A hierarchical cluster analysis was conducted with data from the calibration study, where five different groups were characterized based on demographic and behavioural characteristics.

However, there were no significant differences in the profile of ratings for mice 1 to 9 among these groups. Therefore, all participants' data were jointly analyzed in the following steps of the procedure.

3 KANSEI ENGINEERING STUDY BASED ON AI: KANSAI

This section describes the development and implementation of the KansAI tool, done in four different versions. We will describe these four versions and discuss how each of the versions compares with the results obtained from human ratings, using the validation survey.

3.1 Incremental development of KansAI versions

The KansAI tool was developed through an iterative process, using Anthropic Claude API (claude-sonnet-4-20250514). Claude was selected primarily because a Pro license was provided by Logitech for the purpose of this study, and because Claude was not involved in generating any of the product images used in the evaluation.

Prompts were developed iteratively. The tool was run on all calibration images and the outputs were inspected for failure modes such as product misidentification, insufficient score variation, and unrealistic rating distributions. When a failure mode was identified, the prompt was adjusted to address it, and the process was repeated until outputs consistently met the intended criteria on natural-language instructions.

Each version of KansAI represents the same KansAI tool, but with progressively richer information added to the prompt. The core task, rating scale, Kansei words, and output format were kept constant so the effect of added calibration information could be compared. Table 2 shows the incremental data added in each version.

Table 2. Overview of KansAI versions

Version	Summary	Description
V1	Baseline prompt design	Baseline prompt only. No human-study data, no demographics
V2	Demographic-informed	Adds demographic profiles from the human study: age group, gender, occupation, education level, mouse-use context. No human Kansei ratings are shown.
V3a	KW-rating-informed	Adds human calibration data from 5 calibration mice. Averages of Kansei ratings are given for the 5 mice. Mice are presented with detailed textual visual descriptions
V3b	KW-rating-informed	Same calibration data as V3a, but instead of textual descriptions of the 5 calibration mice, the actual images are embedded and sent as base64 images in the prompt
V4	KW-rating-tuned	Gives human calibration data in the same way as V3a, but including data from the 9 calibration mice (instead of only 5). In addition, it includes some (3) individual ratings from participants as examples

Version V1 is an uncalibrated baseline, testing if Claude could rate product images using Kansei words with only general instructions. Version V2 adds demographic persona information. Version V3a adds averages of human ratings from the calibration study, giving product "images" in a detailed textual form. Version V3b tested the same calibration strategy as V3a but supplied the calibration images directly as base64-encoded visual inputs. It was discontinued due to increased computational cost without observable performance gains. Version V4 expands calibration using all available product images, and giving some individual ratings as examples.

3.2 Comparative Analysis of KansAI Versions

This section presents a comparison of results from the four different versions of the KansAI tool. The indicator chosen to evaluate how well the KansAI tool simulates the ratings of a mouse for each of the 10 kansei words is the mean absolute error. This computation is done in the following manner:

1. Compute the human mean rating ($\bar{H}_{i,k}$) for a given mouse i and Kansei word k.
2. Compute the KansAI mean rating ($\bar{A}_{i,k}$) for the same mouse i and Kansei word k.
3. Take the absolute difference
4. Average those absolute differences across the five validation mice (mice 10 to 14)

So, for one Kansei word:

$$MAE_k = \frac{1}{5}\sum_{i=1}^{5}\left|\bar{H}_{i,k} - \bar{A}_{i,k}\right|$$

Table 3 shows the mean absolute error for each Kansei word in each of the four versions of KansAI.

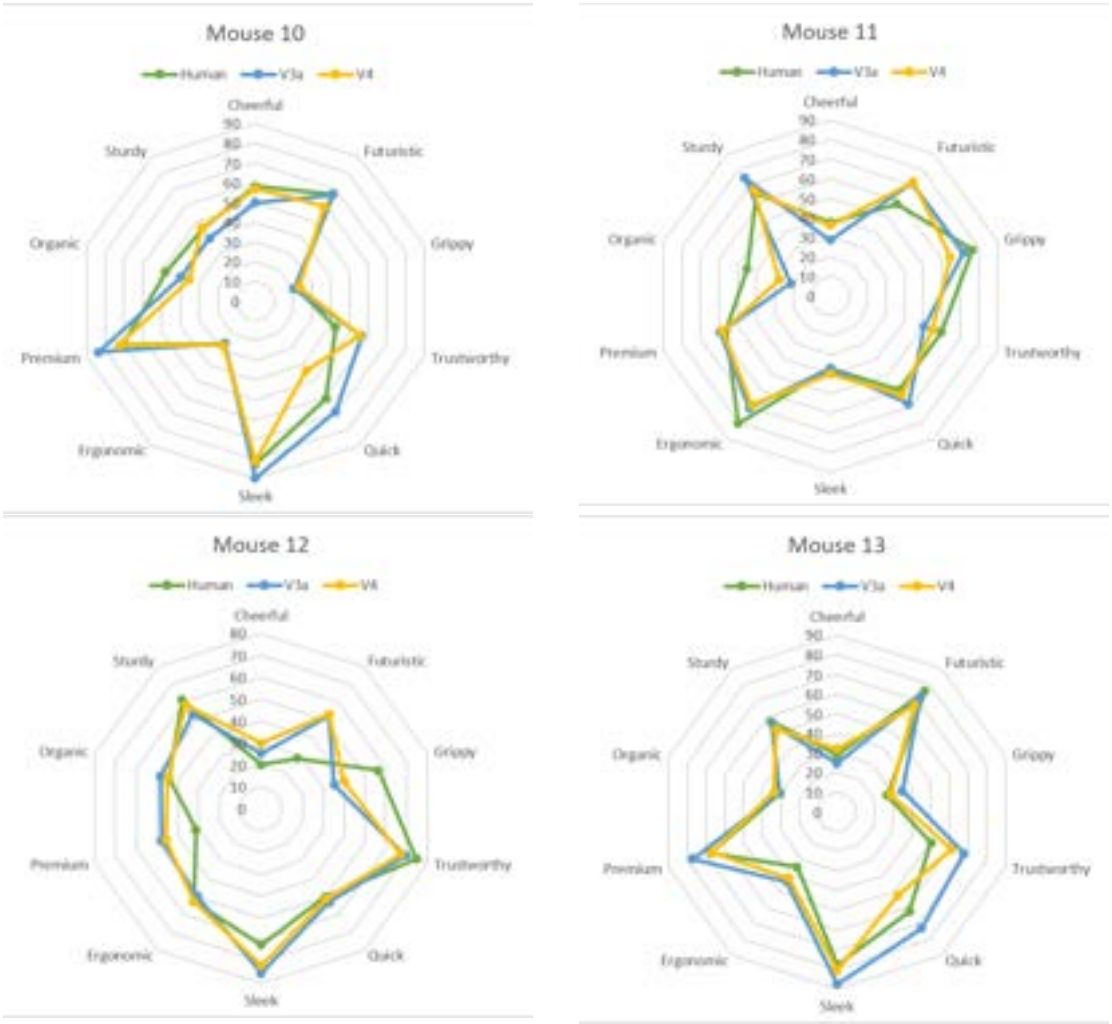
Table 3. Mean absolute error by Kansei word and KansAI version

Kansei word	V1	V2	V3a	V4
Ergonomic	8.09	7.37	6.34	4.46
Sturdy	5.58	5.13	6.65	4.85
Sleek	18.70	18.52	9.57	5.55
Organic	10.35	8.53	7.55	7.05
Premium	17.91	17.31	10.16	7.08
Quick	15.05	12.80	6.92	7.35
Cheerful	21.81	20.36	15.24	9.59
Trustworthy	13.93	12.21	12.79	9.88
Grippy	12.64	13.61	9.03	10.49
Futuristic	17.43	16.90	15.26	17.14
Overall	14.15	13.27	9.95	8.34

Looking at the overall MAE indicator for each version, the largest reduction occurs between V2 and V3a, where the tool was informed with human ratings from the calibration study. Versions V1 and V2 (where no human ratings are informed) have a similar performance. Version V4 is in general slightly better than V3a, but the MAE reduction is modest.

Looking at individual KWs, V4 has the lowest MAE on seven of the ten words. Cheerful and Trustworthy show a large improvement from V3a to V4. Sleek and Premium also improve from V3a to V4, but their main improvement comes from V2 to V3a. Futuristic is a word that shows no improvement over the versions, and is the word where KansAI does a worse job in predicting the rating.

Figure 3 shows radar plots with the profile of average ratings for each mouse and Kansei word for the human ratings (green), and versions V3a (blue) and V4 (yellow). The profiles are, in general, very similar, showing how well the KansAI tool works, especially version V4. However, discrepancies are not negligible in some Kansei words and mice.



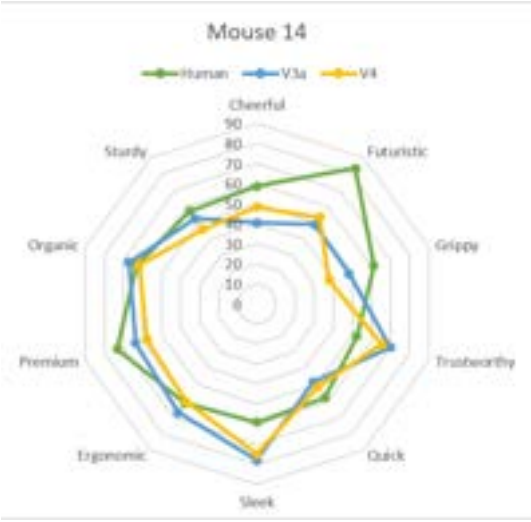


Figure 3. Radar plots showing the profile of ratings given by humans and KansAI (V3a and V4)

3.3 Individual Rating Analysis

As seen in the results from Table 3 and Figure 3, KansAI V3a and V4 are quite good in predicting mean ratings. However, they fail when we want to use them to predict individual ratings. As an example, we can look at Figure 4. For all five mice, the variability of ratings given by KansAI V3a and V4 is much lower than the real variability shown in human ratings. In general, the AI tool correctly locates the level of the response (the average), but, when asked to give individual ratings, it tends to always repeat the same or very similar values, not appropriately capturing the real human variability.

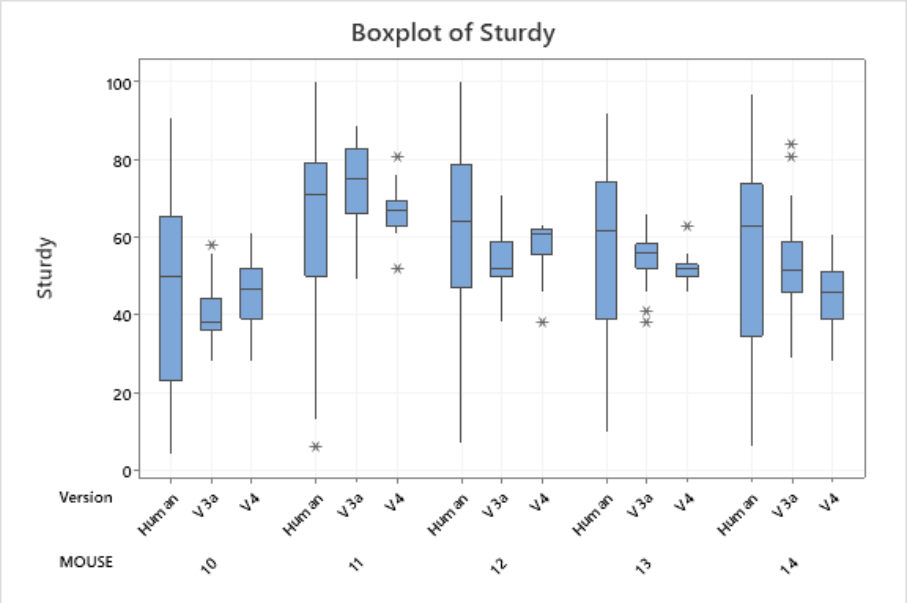


Figure 4. Boxplot of ratings for Sturdy, with Human ratings, and V3a and V4 of KansAI

4 DISCUSSION

The obtained results show that, across the KansAI versions, prediction accuracy generally improved as more human calibration data were incorporated, suggesting that the model could

use examples of human Kansei ratings to better approximate the central tendency of the target group. However, KansAI was not successful at reproducing the variability of individual human responses. This means the tool is better understood as an aggregate estimator.

The variation in accuracy across Kansei words provides important insight into where AI-assisted Kansei evaluation is most promising. Following the distinction proposed by Fenko et al. (2010) and Tavares et al. (2021), the Kansei words can be interpreted as sensory, symbolic, or affective descriptors. KansAI performed best for words with strong visual grounding, such as Ergonomic or Sleek. Although some of these terms involve tactile or use-related qualities, they are still strongly suggested by visible product features such as shape, curvature, and surface texture.

Performance was weaker for words such as Quick and Organic, where the meaning is only partly visible in a static image. These words may depend on assumptions about material, weight, movement, or use experience. The lowest agreement was found for symbolic and affective words such as Futuristic, Cheerful or Trustworthy. These descriptors depend more strongly on cultural expectations, personal interpretation, and emotional response, making them harder to infer from product appearance alone.

Overall, the findings suggest that word category is an important factor in the performance of AI-assisted Kansei evaluation. KansAI appears most reliable when Kansei words are connected to visible product characteristics and less reliable when they depend on subjective, social, or emotional interpretation.

A proof of concept was created and presented to the Head of Industrial Design and the Senior User Experience Designer at Logitech. Both viewed KansAI mainly as a support tool for early-stage design, not as an autonomous replacement for designers. They considered it useful for quick “gut checks” and brainstorming. They also emphasized that the tool should be trustworthy, statistically robust and integrated into existing design workflows if it needs to be really adopted in practice. They said the tool “would need time to prove its usefulness”.

They noted that emotional words are often ambiguous and context-dependent, so Kansei ratings would be more useful if KansAI could explain which visual or physical product features convey each score. Another concern was related to privacy and control of company data.

5 CONCLUSIONS

To sum up, three main conclusions can be drawn:

- KansAI was able to approximate the target group’s general perception of AI-generated computer mouse images at an aggregate level, although it failed to reproduce realistic individual-level variation.
- Calibration with human Kansei rating data substantially improved performance, while demographic information alone produced little improvement.

- Agreement was highest for Kansei words with clear visual grounding, while symbolic and affective words were harder to predict due to their stronger dependence on cultural context, subjective interpretation, and emotional meaning.

REFERENCES

Alcaide-Marzal, J., & Diego-Mas, J. A. (2025). Exploring the use of large language models to build product Kansei semantic spaces. *International Journal of Industrial Ergonomics, 107*, Article 103709. https://doi.org/10.1016/j.ergon.2025.103709

Barros, M., & Ai, Q. (2024). Designing with words: Exploring the integration of text-to-image models in industrial design. *Digital Creativity*, 35(4), 378–391. https://doi.org/10.1080/14626268.2024.2411223

Fenko, A., Otten, J. J., & Schifferstein, H. N. J. (2010). Describing product experience in different languages: The role of sensory modalities. *Journal of Pragmatics*, 42(12), 3314–3327. https://doi.org/10.1016/j.pragma.2010.05.010

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15(1), 3–11. https://doi.org/10.1016/0169-8141(94)00052-5

Norman, D. A. (2003). *Emotional design: Why we love (or hate) everyday things*. Basic Books.

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C. J., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2025, May 20). *Simulating human behavior with AI agents*. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/policy/simulating-human-behavior-with-ai-agents

Schütte, S., Ayas Alikalfa, E., Schütte, R., & Eklund, J. (2006, August 9–11). Developing software tools for Kansei engineering processes: Kansei Engineering Software (KESo) and a design support system based on genetic algorithm [Conference paper]. *9th International Quality Management for Organizational Development (QMOD) Conference*, Liverpool, England.

Schütte, S., Lokman, A. M., Coleman, S., & Marco-Almagro, L. (2023). *Unlocking emotions in design: A comprehensive guide to Kansei engineering* (1st ed.). BoD – Books on Demand.

Tavares, D. R., Canciglieri Junior, O., Guimarães, L. B. de M., & Rudek, M. (2021). A systematic literature review of consumers' cognitive-affective needs in product design from 1999 to 2019. *Frontiers in Neuroergonomics*, 1, Article 617799. https://doi.org/10.3389/fnrgo.2020.617799

SINGLE-PLAYER GAME GENRES AND EMOTIONAL PERFORMANCE: RELATIONSHIPS BETWEEN GAMEPLAY EXPOSURE AND PLAYER EXPERIENCE

Suomiya Bao[1a]

[a] *University of Tsukuba, Japan, bao.suomiya.ka@u.tsukuba.ac.jp*

ABSTRACT

Digital games provide diverse experiential structures that may influence players' psychological experiences and emotional functioning. While previous studies have frequently examined multiplayer interaction in relation to social well-being and motivation, less attention has been given to how different single-player game genres support psychological need satisfaction and player experiences.

Grounded in Self-Determination Theory (SDT), this study investigates the emotional performance of single-player game genres through player experience dimensions including relatedness, competence, autonomy, immersion, and intuitive control. From a Kansei perspective, emotional performance refers to the extent to which interactive systems facilitate desirable affective and psychological experiences during interaction. An online survey was conducted with 122 adult digital game players to examine gameplay exposure across multiple single-player genres and player experiences measured using the Player Experience of Need Satisfaction (PENS) framework. Correlation and multiple regression analyses were performed to examine relationships between gameplay exposure and experiential outcomes.

The results showed that gameplay exposure generally demonstrated positive associations with player experience dimensions. However, after controlling for demographic variables and overlapping gameplay behaviors, most genre effects became non-significant. Fighting game exposure was the only genre that significantly predicted relatedness and immersion. The findings suggest that broad genre classifications alone may have limited explanatory power for emotional performance and that specific experiential characteristics embedded within games may play a more important role in shaping psychological experiences.

Keywords: *Single-player games, Self-Determination Theory, Player Experience, PENS*

[1] Corresponding author.

1 INTRODUCTION

Digital games increasingly function not only as entertainment systems but also as interactive environments capable of influencing users' emotional and psychological experiences. Through narrative structures, challenges, exploration, and player agency, game systems shape motivational and affective responses during interaction. Understanding how specific game design structures facilitate desirable psychological experiences has therefore become an important topic within Kansei Engineering and emotion-oriented interactive system research.

Different game genres provide distinct interaction structures and experiential characteristics that may influence players in different ways. Although video game genre classification remains complex and inconsistent in psychological research (Starosta et al., 2024), previous studies have suggested that game genres influence player engagement, satisfaction, and behavioral outcomes. For example, action and puzzle games have been shown to produce different patterns of player engagement and cognitive responses (Kumar & Singla, 2022), while certain genres such as role-playing and real-time strategy games have been associated with stronger patterns of problematic gameplay behaviors (Eichenbaum et al., 2015, 2017). These findings suggest that game genres may shape psychological experiences through different design characteristics.

Self-Determination Theory (SDT) has frequently been applied to explain psychological experiences within digital games. SDT proposes that human motivation and well-being are supported through satisfaction of three fundamental psychological needs: relatedness, competence, and autonomy (Deci & Ryan, 1985; Ryan & Deci, 2000; Ryan, R. M. & Deci, E. L., 2017). Previous studies have shown that digital games can satisfy these needs through different gameplay structures and interaction mechanisms. Existing research has frequently focused on multiplayer games because direct social interaction naturally facilitates psychological need satisfaction (Kowert et al., 2014). However, single-player games rely on different experiential mechanisms, including narrative engagement, exploration, environmental interaction, and player-controlled progression. Moreover, the increasing use of generative AI in digital environments expands possibilities for game experiences that do not necessarily depend on interactions with other human players. Understanding player experiences in single-player games therefore becomes increasingly important for future game design. Despite the importance of SDT in digital game research, relatively limited studies have examined how different game genres influence player experiences from an SDT perspective. Existing studies suggest that genres may differ in their ability to satisfy psychological needs and shape gameplay experiences (Johnson et al., 2015; Li et al., 2024), yet systematic comparisons across single-player genres remain limited.

Therefore, this study investigates how gameplay exposure across different single-player genres relates to player experience dimensions measured using the SDT framework. By examining relationships between gameplay exposure and psychological experiences, this study contributes to understanding emotional and motivational aspects of game design within Kansei-oriented interactive systems research.

2 OBJECTIVE

The present study investigates emotional performance in single-player digital game genres from the perspective of Self-Determination Theory (SDT). Different game genres provide distinct interaction structures and experiential affordances that may influence players' psychological experiences during gameplay. From a Kansei perspective, emotional performance is defined as the extent to which interactive systems facilitate desirable affective and psychological experiences. In this study, emotional performance is operationalized through player experience dimensions derived from the Player Experience of Need Satisfaction (PENS) framework.

The objectives of this study are:

1. to examine gameplay exposure across different single-player game genres;
2. to investigate relationships between single-player game genre exposure and player experience dimensions;
3. to evaluate emotional performance through psychological need satisfaction and experiential qualities during gameplay.

3 METHODOLOGY

3.1 Participants and Procedure

An online survey was conducted with adult digital game players aged between 18 and 60 years. Participants were recruited through online survey platforms and voluntarily completed the questionnaire. Prior to participation, respondents were screened to confirm that they played digital games at least once per week. Responses with incomplete data were excluded from analysis. The survey consisted of five sections: demographic information, gameplay habits, gameplay exposure measures, player experience measures, and loneliness.

3.2 Survey Design

3.2.1 Demographic Measures

Participants were asked to report demographic information including age, gender.

3.2.2 Gameplay Exposure Measures

Participants reported overall gameplay duration per week and weekly gameplay exposure across nine single-player genres:

- Role-playing games (RPG)
- Story-driven / Adventure games
- Simulation games
- Strategy games
- Puzzle games
- Sandbox games
- Sports games
- Fighting games
- First-person shooter (FPS) games

Participants additionally reported multiplayer gameplay duration as a control variable.

3.2.3 Player Experience Measures

Player experience was assessed using constructs derived from the Player Experience of Need Satisfaction (PENS) framework (Ryan et al., 2006):

- Competence
- Autonomy
- Relatedness
- Immersion
- Intuitive control

Participants responded using a Likert-type agreement scale.

3.2.4 Loneliness

Although the UCLA Loneliness Scale (Russell et al., 1978) was included in the survey for subsequent research, loneliness analyses are beyond the scope of the present paper and are therefore not reported.

3.3 Data Analysis

Descriptive statistics were calculated for demographic characteristics, gameplay exposure, and player experience variables. Internal consistency of all constructs was evaluated using Cronbach's alpha.

Pearson correlation analyses were conducted to examine relationships among gameplay exposure and player experience variables. Multiple regression analyses were subsequently performed to investigate whether gameplay exposure across different single-player genres predicted player experience dimensions while controlling for age, gender, and multiplayer gameplay duration.

4 RESULT

4.1 Descriptive Statistics and Reliability

A total of 122 participants completed the survey. Participants ranged in age from 18 to 60 years (M = 42.16, SD = 11.54). The sample consisted of 59 male participants (48.4%) and 63 female participants (51.6%).

Descriptive statistics for weekly gameplay duration are presented in Table 1. Participants reported an average total gameplay duration of 8.96 hours per week (SD = 7.57). Multiplayer gameplay accounted for an average of 5.78 hours per week (SD = 7.02).

Among single-player game genres, strategy games showed the highest average gameplay duration (M = 4.57, SD = 6.20), followed by shooter games (M = 4.21, SD = 6.35), adventure games (M = 4.14, SD = 5.71), and puzzle games (M = 4.04, SD = 5.56). Relatively lower average gameplay durations were observed for fighting games (M = 3.28, SD = 4.78), sandbox games (M = 3.59, SD = 5.89), sports games (M = 3.68, SD = 5.98), and role-playing games (M = 3.68, SD = 5.71). Large

standard deviations across genres suggest considerable variation in gameplay preferences among participants.

Table 1. Weekly Gameplay Duration by Genre (Hours per Week)

Genre	Mean (hours/week)	SD
Total	8.96	7.57
RPG	3.68	5.71
Adventure	4.14	5.71
Simulation	3.74	5.68
Strategy	4.57	6.20
Puzzle	4.04	5.56
Sandbox	3.59	5.89
Sports	3.68	5.98
Fighting	3.28	4.78
Shooter	4.21	6.35
Multi-player	5.78	7.02

Reliability analyses were conducted to evaluate the internal consistency of the psychological measures used in this study (Table 2). Overall, the scales demonstrated acceptable to good reliability, with Cronbach's alpha values ranging from .82 to .91, indicating satisfactory internal consistency across all constructs.

Initial reliability analysis of the relatedness dimension revealed relatively lower internal consistency. Examination of item-level statistics indicated that one item referring specifically to closeness with non-player characters (NPCs) showed poor consistency with the remaining relatedness items and substantially reduced the overall reliability of the construct. Conceptually, this item was considered less suitable for the present study because not all single-player game genres involve meaningful NPC interactions. Therefore, the item was removed from subsequent analyses. The final relatedness construct consisted of two items and demonstrated high reliability ($\alpha = .91$).

Among the PENS dimensions, relatedness demonstrated the highest reliability ($\alpha = .91$), followed by immersion ($\alpha = .87$), intuitive control ($\alpha = .84$), autonomy ($\alpha = .83$), and competence ($\alpha = .82$).

Regarding descriptive statistics, participants reported relatively high levels of competence ($M = 5.60$, $SD = 1.17$), autonomy ($M = 5.66$, $SD = 1.24$), and intuitive control ($M = 5.48$, $SD = 1.24$) during gameplay experiences. Relatedness showed comparatively lower ratings ($M = 4.95$, $SD = 1.85$), suggesting greater variability in perceived relational experiences across single-player games. Immersion demonstrated a moderate to high level ($M = 4.99$, $SD = 1.25$).

Table 2. Means, Standard Deviations, and Internal Consistency of Measured Constructs

Construct	items	mean	SD	Cronbach's α
PENS_ competence	3	5.60	1.17	0.82
PENS_ autonomy	3	5.66	1.24	0.83
PENS_ relatedness	2	4.95	1.85	0.91
PENS_ immersion	9	4.99	1.25	0.87
PENS_ Intuitive control	3	5.48	1.24	0.84

Note: One relatedness item ("I don't feel close to other NPCs") was removed due to low internal consistency and conceptual incompatibility across game genres.

4.2 Correlation Analysis

Pearson correlation analysis was conducted to examine relationships among gameplay exposure, player experience dimensions. The results are presented in Figure 1.

Strong positive correlations were observed among gameplay durations across different game genres ($r = .44–.76$, $p < .001$), indicating that participants who spent more time on one genre also tended to engage with multiple other genres. Multiplayer gameplay duration also showed significant positive correlations with most single-player genres ($r = .44–.71$, $p < .001$).

Regarding player experience dimensions, gameplay exposure generally demonstrated weak to moderate positive relationships with PENS measures. Relatedness showed significant positive associations with most game genres, including role-playing games ($r = .34$, $p < .001$), strategy games ($r = .31$, $p < .001$), sports games ($r = .30$, $p < .01$), and fighting games ($r = .40$, $p < .001$). Similar patterns were observed for competence, autonomy, immersion, and intuitive control, suggesting that greater gameplay engagement was generally associated with higher levels of perceived psychological need satisfaction and positive player experiences.

Strong interrelationships were also observed among the PENS dimensions themselves. Relatedness demonstrated particularly strong correlations with immersion ($r = .86$, $p < .001$), competence ($r = .62$, $p < .001$), autonomy ($r = .64$, $p < .001$), and intuitive control ($r = .62$, $p < .001$). Competence and autonomy were also strongly correlated ($r = .85$, $p < .001$), indicating substantial overlap among experiential dimensions during gameplay.

4.3 Effects of Game Genre Exposure on Player Experience

Multiple regression analyses were conducted to examine whether gameplay exposure across different single-player genres predicted player experience dimensions while controlling for age,

gender, and multiplayer gameplay exposure. Separate models were estimated for competence, autonomy, relatedness, immersion, and intuitive control. Figure 2 shows the results.

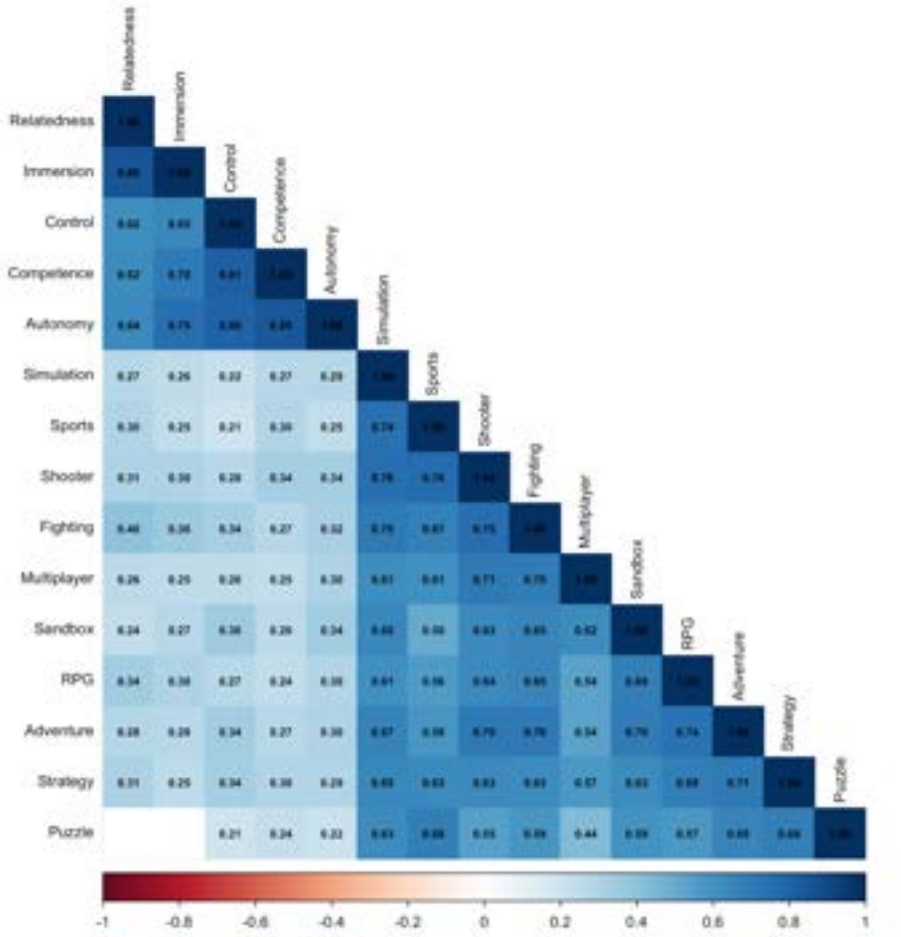


Figure 1. Pearson correlation matrix among gameplay exposure and player experience dimensions. Only statistically significant correlations ($p < .05$) are displayed

Overall, genre exposure demonstrated limited predictive effects across the five player experience dimensions. Among the genre variables, fighting game exposure emerged as the only significant predictor of multiple experiential outcomes. Specifically, fighting game exposure positively predicted relatedness ($\beta = .41$, $p = .007$) and immersion ($\beta = .32$, $p = .038$). No significant effects were observed for role-playing, adventure, simulation, strategy, puzzle, sandbox, sports, shooter, or multiplayer gameplay exposure. Gender also demonstrated significant effects across several experiential dimensions. Male participants reported significantly higher relatedness ($\beta = .46$, $p = .015$), competence ($\beta = .48$, $p = .017$), and immersion ($\beta = .45$, $p = .022$) compared with female participants. Age did not significantly predict any player experience dimensions.

The regression models explained between 16.9% and 25.9% of variance across player experience dimensions ($R^2 = .169–.259$). Multicollinearity diagnostics indicated acceptable variance inflation factor values across predictors (VIF = 1.17–4.35), suggesting that predictor overlap did not substantially affect model stability.

These findings suggest that broad genre categories alone may have limited influence on psychological gameplay experiences. Instead, emotional performance may be influenced more strongly by specific experiential characteristics embedded within games than by genre classification itself.

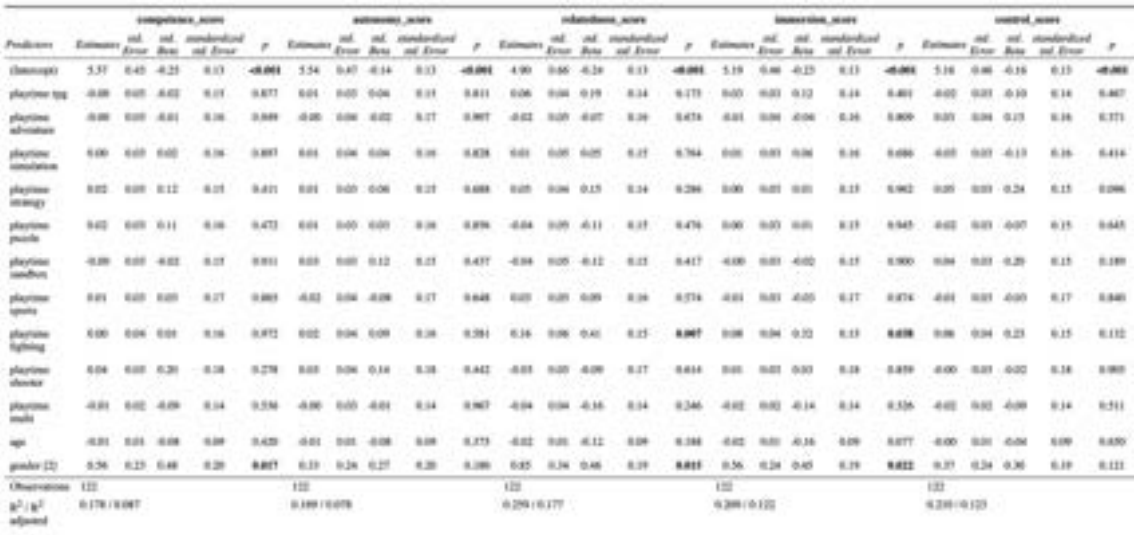


Figure 2. Multiple regression coefficients predicting player experience dimensions (relatedness, competence, autonomy, immersion, and intuitive control) from single-player gameplay exposure while controlling for age, gender (1 = female, 2 = male), and multiplayer gameplay duration. Significant effects are indicated by asterisks ($p < .05$, *$p < .01$, $p < .001$).

5 DISCUSSION

The present study investigated the emotional performance of single-player digital game genres through the lens of Self-Determination Theory and Kansei-oriented perspectives on interactive experience. Specifically, the study examined whether gameplay exposure across different single-player genres was associated with psychological need satisfaction and player experience dimensions, including relatedness, competence, autonomy, immersion, and intuitive control.

Contrary to the initial expectation, broad game genre categories demonstrated relatively limited predictive effects on player experience dimensions. Although positive associations between gameplay exposure and psychological experience were observed in the correlation analysis, most genre effects disappeared after controlling for overlapping gameplay behaviors and demographic variables in the regression models. These findings suggest that broad genre classifications alone may not sufficiently explain variations in emotional performance. One possible explanation is that genre categories represent relatively coarse descriptions of game characteristics and may contain substantial internal variability. For example, role-playing games may differ considerably in narrative depth, social interaction opportunities, character attachment, or exploration mechanics. Similarly, adventure and simulation games can provide highly diverse emotional and experiential structures despite belonging to the same genre category. Consequently, psychological experiences may be shaped more strongly by specific design elements and experiential affordances than by genre labels themselves.

Interestingly, fighting game exposure positively predicted both relatedness and immersion. This finding differs from the common assumption that relatedness in games primarily emerges from explicit social interaction or narrative engagement. A possible interpretation is that repeated engagement with recognizable characters, avatar identification, mastery processes, and emotional attachment to game systems may contribute to a sense of connection even within traditionally less socially oriented game genres. Because the present study focused on gameplay exposure rather than specific interaction mechanisms, additional investigation is needed to clarify the underlying processes responsible for this relationship. Gender differences were also observed, with male participants reporting higher relatedness, competence, and immersion scores. However, gender was not a primary focus of this study, and further investigation is required before drawing broader conclusions regarding gender-related differences in gameplay experiences.

From a Kansei Engineering perspective, the findings suggest that emotional performance may not be adequately represented through broad content classifications such as game genres. Instead, experiential qualities such as narrative engagement, perceived agency, emotional attachment, and interaction characteristics may provide more direct indicators of affective performance within digital environments. This perspective aligns with the view that emotional outcomes emerge from interactions between users and specific design structures rather than from categorical system labels alone.

The present study has several limitations. First, the use of self-reported gameplay exposure may introduce recall bias and estimation inaccuracies. Second, the cross-sectional design limits causal interpretation of the observed relationships. Third, the current study focused on general genre exposure and did not examine specific game titles or detailed design characteristics. Future studies should investigate more fine-grained experiential variables and explore potential relationships between player experience dimensions and psychological outcomes such as loneliness and well-being.

6 CONCLUSION

This study examined the emotional performance of single-player digital game genres by investigating relationships between gameplay exposure and player experience dimensions grounded in Self-Determination Theory. Although gameplay exposure across genres generally showed positive associations with player experience variables, broad genre classifications demonstrated limited predictive effects after controlling for demographic characteristics and overlapping gameplay behaviors. Among the examined genres, only fighting game exposure significantly predicted relatedness and immersion, suggesting that certain experiential characteristics embedded within games may contribute to psychological experiences independently of traditional assumptions regarding social interaction or narrative structure.

Overall, the findings suggest that emotional performance in digital games may be better understood through specific experiential and design characteristics rather than through broad genre categories alone. These findings contribute to Kansei-oriented perspectives on interactive

systems by highlighting the importance of examining how design structures shape psychological experiences within digital environments.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP26K21362.

REFERENCES

Deci, E. L., & Ryan, R. M. (1985). Intrinsic Motivation and Self-Determination in Human Behavior. Springer US. https://doi.org/10.1007/978-1-4899-2271-7

Eichenbaum, A., Kattner, F., Bradford, D., Gentile, D. A., Choo, H., Chen, V. H. H., Khoo, A., & Green, C. S. (2017). The Role of Game Genres and the Development of Internet Gaming Disorder in School- Aged Children. Journal of Addictive Behaviors,Therapy & Rehabilitation. https://doi.org/10.4172/2324-9005.1000141

Eichenbaum, A., Kattner, F., Bradford, D., Gentile, D. A., & Green, C. S. (2015). Role-Playing and Real-Time Strategy Games Associated with Greater Probability of Internet Gaming Disorder. Cyberpsychology, Behavior, and Social Networking, 18(8), 480–485. https://doi.org/10.1089/cyber.2015.0092

Johnson, D., Nacke, L. E., & Wyeth, P. (2015). All about that Base: Differing Player Experiences in Video Game Genres and the Unique Case of MOBA Games. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, 2265–2274. https://doi.org/10.1145/2702123.2702447

Kowert, R., Domahidi, E., Festl, R., & Quandt, T. (2014). Social gaming, lonely life? The impact of digital game play on adolescents' social circles. Computers in Human Behavior, 36, 385–390. https://doi.org/10.1016/j.chb.2014.04.003

Kumar, R., & Singla, R. (2022). Attention guided EEG based Game Genre Classification. 2022 2nd International Conference on Intelligent Technologies (CONIT), 1–6. https://doi.org/10.1109/CONIT55038.2022.9847911

Li, X., Lenarduzzi, V., & Taibi, D. (2024). A Data-Driven Analysis of Player Personalities for Different Game Genres. Proceedings of the 1st ACM International Workshop on Foundations of Applied Software Engineering for Games, FaSE4Games 2024, 20–25. https://doi.org/10.1145/3663532.3664468

Russell, D., Peplau, L. A., & Ferguson, M. L. (1978). Developing a measure of loneliness. Journal of Personality Assessment, 42(3), 290–294. https://doi.org/10.1207/s15327752jpa4203_11

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. Guilford Press.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motivation and Emotion, 30(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8

Starosta, J., Kiszka, P., Szyszka, P. D., Starzec, S., & Strojny, P. (2024). The tangled ways to classify games: A systematic review of how games are classified in psychological research. PLOS ONE, 19(6), e0299819. https://doi.org/10.1371/journal.pone.0299819

THE EFFECT OF DIGITAL SHEET MUSIC INTERFACES ON BEGINNERS' AUTONOMOUS PRACTICE, FLOW AND COGNITION

Wen-Hsien Kuo[a], Meng-Cong Zheng[1b]

[a] *Department of Industrial Design, National Taipei University of Technology, Taipei, Taiwan*
hsienxx@gmail.com
[b] *Department of Industrial Design, National Taipei University of Technology, Taipei, Taiwan*
zmcdesign@gmail.com

ABSTRACT

With hardware upgrades and the spread of technology, piano learning has shifted from traditional paper scores to digital interfaces offering dynamic guidance and real-time feedback. However, how information presentation in digital sheet music influences beginners' flow experience and cognitive load during self-practice remains underexplored. This study investigates the effects of interface factors (notation format and animation type) on beginners' practice performance and psychological experience. A mixed-method design was adopted using a 2 (notation format: vertical, horizontal) × 2 (animation format: animated image, hand video) between-subject experiment with 40 novices. Participants practiced seven progressively difficult pieces. Post-test measures included the Flow Experience Scale, Cognitive Load Scale, System Usability Scale, and semi-structured interviews. Results showed three main findings: (1) Performance strategy and flow: Participants using sheet music during playing achieved significantly higher flow and concentration than those relying solely on video (p=.004). (2) Interface interaction: The "vertical notation + hand video (Ab)" combination yielded the best flow (M=4.08) and lowest mental effort (M=3.10), indicating optimized element pairing enhances learning immersion. (3) Subjective preference: Although Ab performed best, participants favored Bb (horizontal notation + hand video), likely due to gaze-switching distance. Interview analysis further revealed expectations for predictive cues and hierarchical practice designs to reduce notation-reading errors. These findings suggest specific combinations of notation and animation formats can enhance flow and reduce cognitive load. The study provides concrete directions for digital sheet music design to lower beginners' learning barriers and support immersive piano practice.

Keywords: *User experience, System Usability, Cognitive Load Theory, Optimal Experience*

[1] Corresponding author. 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608 Taiwan, zmcdesign@gmail.com

1 INTRODUCTION

The acquisition of piano performance is a highly labor-intensive activity that requires long-term commitment to structured "deliberate practice" (Ericsson et al., 1993). For beginners, maintaining stable practice quality is the key to developing piano literacy. However, research indicates that without systematic "formal practice," beginners often face bottlenecks of learning stagnation (Bonneville-Roussy & Bouffard, 2015). With the development of digital hardware and multimedia technology, particularly in the post-pandemic era, score formats have shifted from traditional paper to digital interfaces featuring animated guidance, split screens, and real-time evaluation functions (Wang, 2022; Wu, 2023), providing immediate reference resources for autonomous practice. Although digital tools can enhance motivation, improper presentation of interface information may also trigger cognitive overload and hinder more refined musical expression (Rogers et al., 2014). In learning contexts, users' "flow experience" and "cognitive load" are key indicators for measuring learning effectiveness (Engeser & Schiepe-Tiska, 2012; Sweller, 1998). Flow refers to the state of immersion produced when an individual is fully concentrated on a certain activity, which can enhance engagement and a sense of control (Csikszentmihalyi, 2014). On the other hand, imperfect instructional design generates "extraneous cognitive load," consuming the learner's limited working memory capacity (Sweller, 1998). Although digital media aims to provide motivating interfaces, excessive redundant information may lead to cognitive overload in beginners(Klepsch et al., 2017).

Current digital score apps (such as Flowkey, Skoove, etc.) employ diverse visual cues, such as dynamic guidance, fingering video demonstrations, and color-coded prompts (Molero et al., 2021; Wu, 2023). However, the impact of different interface presentation forms—such as the vertical or horizontal layout of the score, and the visualization differences in keyboard dynamic images (animated images versus hand recordings)—on beginners' score-reading behavior and psychological perceptions remains inconclusive. This study aims to investigate the specific effects of different digital interface forms on beginners' practice effectiveness, flow state, and learning cognition.

To achieve the aforementioned objectives, this study employs an experimental design: a 2 (Score Format: Vertical/Horizontal) x 2 (Animation Form: Digital Animation/Hand Recording) two-factor design to evaluate beginners' experiences with different interface combinations. By analyzing the results, interface optimization design and verification will be conducted, aiming to provide specific reference standards for future digital score design.

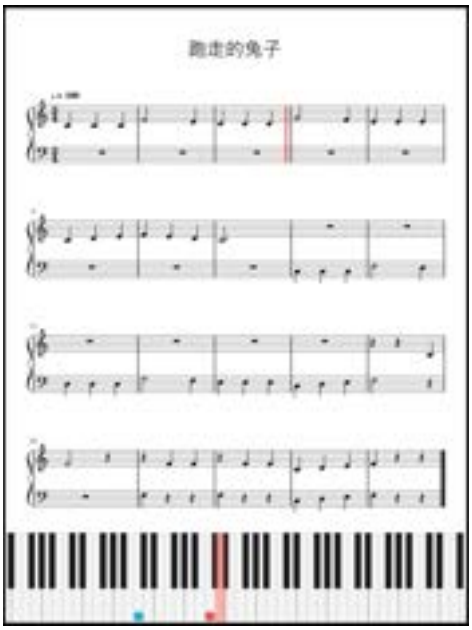
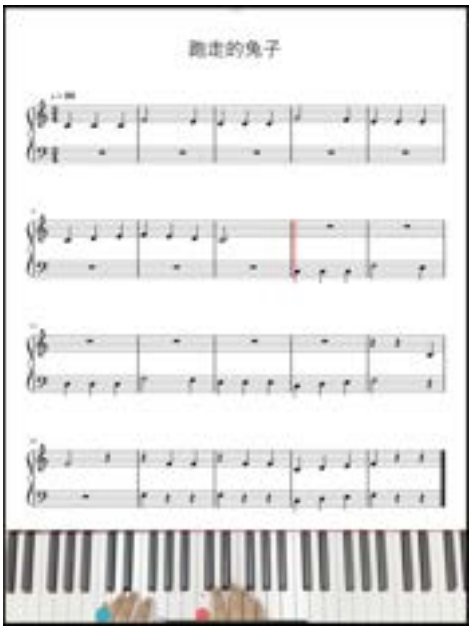
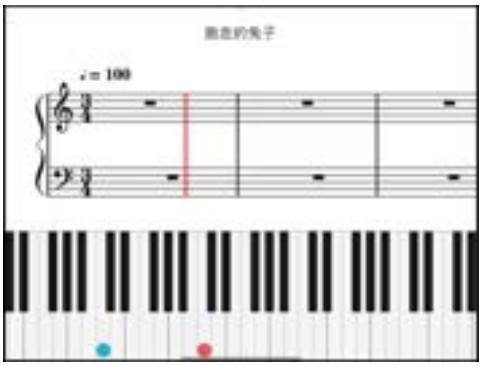
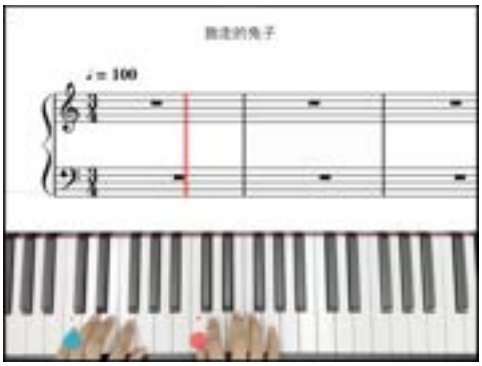
2 METHOD

This study adopts a Mixed Method design for the experiment: by analyzing existing digital musical score software on the market, key design factors affecting learning are clarified.

2.1 Experiment on Digital Musical Score Practice Simulation

The study adopts a 2x2 between-subjects factorial experimental design (Between-subjects Factorial Design). Independent variables are defined as follows: Variable 1: Musical Score Presentation Format (A) Vertical Score: Simulating traditional paper layout, presenting 3 to 4 lines of musical score simultaneously. (B) Horizontal Score: Adopting a single-line horizontal scrolling presentation. Variable 2: Instructional Animation Format (a) Dynamic Image: Using synchronized digital color blocks and virtual keyboard prompts. (b) Hand Recording: Providing performance videos of a real pianists, supplemented with color-coded fingering. Based on the above variables, four sets of experimental samples were combined: Aa (Vertical + Animation), Ab (Vertical + Recording), Ba (Horizontal + Animation), and Bb (Horizontal + Recording). All experimental samples were developed using Adobe After Effects for simulation to ensure precise synchronization of audio, score scrolling, and fingering images.

Table 1. Visual Presentation of Experimental Samples

	
Aa: Vertical Score / Animated Graphics	**Ab**: Vertical Score / Hand Recording
	
Ba: Horizontal Score / Animated Graphics	**Bb**: Horizontal Score / Hand Recording

This experiment recruited 40 participants with completely no piano learning experience (equally distributed into four groups, n=10 per group), to eliminate interference from prior knowledge. Experimental equipment utilized a Yamaha P-525 digital piano and an iPad Pro 12.9-inch screen as the display interface. The experimental tasks were designed in collaboration with professional piano teachers, selecting 7 musical pieces with a difficulty gradient, including 3 right-hand fingering, 3 left-hand fingering, and 1 two-hand ensemble task (such as "Mary Had a Little Lamb," "The Runaway Rabbit," etc.) The procedure was as follows: (1) Preliminary preparation: Conduct 3 to 5 minutes of basic music theory explanation (such as note durations and finger-to-key mapping), and basic scale practice to construct preliminary concepts; (2) Formal experiment: Each musical piece was practiced in sequence with "slow speed (80 BPM) twice" and "normal speed (100 BPM) once," without providing human guidance during the process. After completing the tasks, participants were required to complete the following scales:

1. Flow Experience Scale: Assessing challenge balance, concentration, and immersion (Kiili, 2006; Hou & Chou, 2012).
2. Cognitive Load Scale: Measuring mental load and mental effort levels (Chu et al., 2010; Hwang et al., 2013).
3. System Usability Scale (SUS): Assessing the operability of the interface. Finally, participants were required to perform preference ranking for the four group samples and undergo a semi-structured interview to deeply explore their gaze-switching behavior and reasons for visual dependence (Bangor, 2009).

3 RESULT

This study conducted digital musical score tests of 4 samples on 40 participants without piano learning experience through experiments, and performed analysis from four aspects: subject characteristics, playing strategies, quantitative scales, and interview feedback, to clarify the impact of interface combination forms on learner experience.

3.1 Analysis of Participant Characteristics and Playing Strategies

There were a total of 40 experimental participants (17 males, 23 females), with ages distributed between 18 and 29 years old, and all had no piano learning experience. Analysis of participant performance across seven musical pieces revealed that even as beginners, they exhibited significant differences in "playing strategy differences" when facing playing tasks. Based on behavioral self-assessments, participants were categorized into two primary groups: (1) Non-score using group (50%): Completely relied on the keyboard animation below (100% video), and did not pay attention to the information on the staff. (2) Score using group (50%): Referred to the blocks of the staff to varying degrees during the performance process (30-70% score), attempting to establish a correspondence between fingering and scales.

3.2 Quantitative Impact of Performance Strategies on Learning Experience

To investigate the impact of whether beginners possess "performance strategies" on the learning experience, this study utilized a One-Way Analysis of Variance (One-Way ANOVA) for

testing: Results showed that participants who used scores were significantly higher in total flow scores than those who did not use scores (p=.004). This difference was simultaneously reflected in the two sub-constructs of "flow antecedents" (p=.040) and "flow experience" (p=.006). The data illustrate that when learners actively allocate attention to the musical score, they can establish clearer learning goals and feedback mechanisms, thereby generating high concentration and a positive sense of immersion.

Table 2. One-Way ANOVA of Flow Experience by Performance Strategy

Measurement Variables	Score-Using Group (n=20) *M (SD)*	Non-Score-Using Group (n=20) *M (SD)*	*F*	*p*
Flow Antecedents	3.50 (0.54)	3.13 (0.55)	4.532	.040*
Flow Experience	3.97 (0.52)	3.50 (0.49)	8.57	.006**
Total Flow Score	3.75 (0.42)	3.33 (0.44)	9.51	.004**

* p < .05, ** p < .01

The computational results for cognitive load indicated that there were no significant differences between the two groups in mental load, mental effort, or total cognitive load scores. However, in terms of mean trends, the group without score-reading habits (M = 4.37) was slightly higher than the group with habits (M = 3.71), indicating that not looking at the score may lead to higher mental load. In terms of System Usability Scale (SUS), the evaluations for the two groups were 74 points and 69 points respectively, both reaching the "Good" level, confirming that the interface design possesses sufficient versatility and that the operational threshold will not change significantly due to score-reading behavior.

3.3 Interaction Analysis of Interface Combination Forms

In addition to performance strategies, this study also conducted a specific analysis based on the 2 (Score Format: Vertical/Horizontal) x 2 (Animation Style: Animated Graphics/Hand Recording) two-factor experimental design, as follows:

Through Two-Way Analysis of Variance (Two-Way ANOVA), it was found that "score format" and "animation format" have significant interactions in flow experience (p = .020), total flow score (p = .019) (Table 3), and cognitive mental effort (p = .009) (Table 4). Data show that the "vertical score + hand recording (Ab)" combination can create the highest flow experience (M = 4.08) and reduce mental effort to the minimum (M = 3.10). The "horizontal score + hand recording (Bb)" was relatively poor in its performance in these two items (Flow experience: M = 3.46, mental effort: M = 3.97). This indicates that when paired with hand recording (b), the vertical score (A) is the best combination.

3.4 Subjective Preferences and Interview Findings

At the end of the experiment, participants were invited to rank the preferences of the four samples and undergo qualitative interviews to understand the operational context behind the participants. According to the Friedman test, the highest subjective preference among all participants was Bb (horizontal score + hand recording) (p < .001). However, if subdivided by strategy groups, the "Score Group" most preferred Ab (M=1.95), while the "No-score Group"

most preferred Bb (M=1.55). This result illustrates that the realism provided by hand recording is attractive to all beginners, but in terms of score layout, those with score-using strategies are more inclined toward a vertical layout.

Table 3. Two-Way ANOVA of Flow Experience Across Four Samples

Dependent Variable	Source of Variation	SS	df	MS	F	p	ηp2
Flow Antecedents	Score Format	0.01	1	0.01	0.04	.845	.001
	Animation Style	0.18	1	0.18	0.57	.454	.016
	Interaction	0.87	1	0.87	2.74	.107	.071
	Error	11.44	36	0.32			
Flow Experience	Score Format	0.47	1	0.47	1.74	.196	.046
	Animation Style	0.03	1	0.03	0.13	.725	.003
	Interaction	1.6	1	1.6	5.92	0.02*	.141
	Error	9.74	36	0.27			
Total Score	Score Format	0.1	1	0.1	0.51	.480	.014
	Animation Style	0.09	1	0.09	0.42	.520	.012
	Interaction	1.24	1	1.24	6.04	0.019*	.144
	Error	7.39	36	0.21			

* p < .05, ** p < .01

Table 4. Two-Way ANOVA of Cognitive Load Across Four Samples

Dependent Variable	Source of Variation	SS	df	MS	F	p	ηp2
Mental Load	Score Format	0.2	1	0.2	0.12	.730	.003
	Animation Style	1.3	1	1.3	0.8	.378	.022
	Interaction	2.12	1	2.12	1.3	.261	.035
	Error	58.46	36	1.62			
Mental Effort	Score Format	0.03	1	0.03	0.02	.881	.001
	Animation Style	1.47	1	1.47	1.34	.256	.036
	Interaction	8.4	1	8.4	7.63	.009**	.175
	Error	39.63	36	1.1			
Total Score	Score Format	0.11	1	0.11	0.09	.767	.002
	Animation Style	1.36	1	1.36	1.07	.307	.029
	Interaction	3.98	1	3.98	3.15	.084	.080
	Error	45.57	36	1.27			

* p < .05, ** p < .01

Table 5. Friedman Test for Overall Preference Ranking: Non-parametric Comparison of Multiple Dependent Samples

Sample Combination	*n*	*M*	*SD*	Mean Rank	Preference Rank
Bb	40	1.78	0.97	1.78	1
Aa	40	2.15	0.98	2.15	2
Ba	40	2.78	1	2.78	3
Ab	40	3.3	0.91	3.3	4

Results of the Friedman test：(3) = 32.73,** $p < .001$

Table 6. Friedman Test for Preference Ranking Grouped by Performance Strategy: Non-parametric Comparison of Multiple Dependent Samples

Sample Combination	Score-Using (n=20)			Non-Score-Using (n=20)		
	M	*SD*	Preference Rank	*M*	*SD*	Preference Rank
Ab	1.95	0.95	1	2.35	0.99	2
Bb	2.00	1.12	2	1.55	0.76	1
Ba	3.00	0.92	3	2.55	1.05	3
Aa	3.05	1.05	4	3.55	0.69	4

Friedman test results for the Score-Using Group:$\chi^2(3) = 13.26$, $^{**}p = .004$
Friedman test results for the Non-Score-Using Group:$\chi^2(3) = 24.36$, $^{**}p < .001$

Through the organization of participant interview content, the following core findings were summarized: Beginners exhibit an extreme reliance on interface guidance elements (such as red lines and color blocks) during the practice process and regard real hand recordings as a core tool for alleviating the pressure of scale recognition.

For participants with a score-reading inclination, their learning behavior follows a hierarchical path from imitating fingering movements to establishing sensory melodic memory, and finally to verifying musical staff symbols, indicating that digital interfaces can effectively serve as a transitional bridge before entering the domain of traditional score reading. However, due to the perceptual latency caused by the synchronized display of prompts and music in existing systems, participants are often unable to react in time, leading to a strong demand for visual anticipation, such as displaying the next key position in advance.

Especially when facing high-demand passages involving two-hand coordination or wide-scale movements, the sudden increase in cognitive load frequently caused participants to exceed their processing threshold, resulting in a complete abandonment of score reading and regression to a passive, video-following mode. Notably, melodic familiarity played a significant moderating role in this process: prior memory of the melody effectively reduced visual dependence on the interface, transforming the system's function from a dominant navigation role into a supportive reference tool.

4 DISCUSSION

This study explores the impact of digital music score interfaces on beginners' playing practice through experiments and analysis. The following is a deep discussion on playing strategies, interface fitness, and the gap between subjective and objective evaluation.

4.1 The Key Role of Playing Strategies on Learning Immersion

Research results show that whether participants adopt the strategy of "reading music scores" significantly affects their perceived level of flow. Although it is highly challenging for those without piano learning experience to read scores and play simultaneously, statistical data shows that the score-using group outperformed the non-score-using group in total flow score (p=.004), flow antecedents (p=.040), and flow experience (p=.006). This echoes Csikszentmihalyi's definition of flow antecedents: clear goals and immediate feedback can guide the focus of attention (Csikszentmihalyi, 2014). When beginners attempt to establish logical connections between scales and fingerings, music scores provide structured task guidance, which in turn transforms into higher concentration and positive immersion.

4.2 Interaction of Interface Elements and Patterns of Fitness

This study confirmed through two-way ANOVA that the quality of digital score interfaces is not determined by a single element, but depends on the interactive pairing of score format and animation format. Best Performance Combination (Ab): Experimental data show that "vertical score + hand recording" performed best in enhancing flow experience (M=4.08) and reducing cognitive mental effort (M=3.10).

The interview results further confirm that the vertical score can provide better information continuity and preview space, while hand recording can effectively reduce the conversion pressure between score scales and actual piano key mapping, reducing the burden of gaze-switching. The research results also point out significant compatibility differences between interface elements; specifically, the vertical score is suitable for pairing with concrete hand recording, while the horizontal score is more compatible with geometric animated graphics. This illustrates that under different visual layouts, there are differences in the brain's efficiency in processing concrete actions and abstract symbols; therefore, future design adjustments should deeply consider the gaze-switching distance and the complexity of information processing.

4.3 Analysis of the Gap Between Subjective Preference and Objective Performance

Another noteworthy finding is the gap between participants' "preference" and "performance." Although the Ab combination performed best in flow and cognitive load indicators, according to the Friedman test, the favorite of all participants was Bb (horizontal score + hand recording). This may be because the horizontal layout makes the proportions of the score and animation blocks similar, with short visual movement distances, giving beginners an intuitive feeling of "easier concentration." However, because music and images appear simultaneously, even with high focus on the screen, delays still occur, causing participants to fall into a state of passive following.

4.4 Optimization Suggestions for Digital Music Scores

Combining the results of quantitative data and qualitative interviews, this study proposes the following optimization directions for digital sheet music interfaces. First, to alleviate the perceptual latency and reaction discrepancy caused by synchronized display, a visual anticipation mechanism should be introduced to pre-announce subsequent note positions in advance, thereby reducing the reaction pressure on participants.

Second, according to the participants' three-stage behavioral strategy of "imitation, memory, and verification," a hierarchical practice mode should be established. For example, fingering video guidance should be strengthened in the initial stage to establish muscle memory, while video navigation should be gradually weakened and staff symbols strengthened in the later stages to assist users in transitioning from system dependence to the goal of independent score-reading and performance.

Finally, to reduce the cognitive dissonance and visual fatigue caused by multiple dynamic elements, the interface layout should be optimized to reduce dynamic interference. Especially when performing high-difficulty tasks such as hands-together ensembles, background information should be simplified or clearer early-warning mechanisms should be provided to avoid the psychological frustration caused by rhythmic breakdown in beginners.

5 CONCLUSION

This study reveals the profound impact of digital music score interface design on piano beginners' autonomous practice. The research confirms that guiding users to actively read scores during the practice process, rather than only passively imitating video prompts, can significantly enhance immersion and concentration levels during learning. In addition, a gap exists between beginners' subjective preferences and actual operational performance, illustrating that interface design should transcend simple visual intuition and prioritize resolving obstacles such as perceptual delay that affect rhythmic accuracy. Future design optimization should introduce guidance functions with "anticipation mechanisms" and combine them with the "imitation—memory—verification" hierarchical teaching path. Through this progressive auxiliary framework, learners can be assisted in transitioning smoothly from complete reliance on navigation to independent performance, achieving a more ideal autonomous learning experience.

REFERENCE

Bonneville-Roussy, A., & Bouffard, T. (2015). When quantity is not enough: Disentangling the roles of practice time, self-regulation and deliberate practice in musical achievement. *Psychology of Music*, *43*(5), 686–704. https://doi.org/10.1177/0305735614534910

Csikszentmihalyi, M. (2014). *Applications of Flow in Human Development and Education: The Collected Works of Mihaly Csikszentmihalyi*. Springer Netherlands. https://doi.org/10.1007/978-94-017-9094-9

Engeser, S., & Schiepe-Tiska, A. (2012). Historical Lines and an Overview of Current Research on Flow. In S. Engeser (Ed.), *Advances in Flow Research* (pp. 1–22). Springer New York. https://doi.org/10.1007/978-1-4614-2359-1_1

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). *The Role of Deliberate Practice in the Acquisition of Expert Performance*.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Molero, D., Schez-Sobrino, S., Vallejo, D., Glez-Morcillo, C., & Albusac, J. (2021). A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and Applications*, *80*(1), 165–186. https://doi.org/10.1007/s11042-020-09678-9

Rogers, K., Röhlig, A., Weing, M., Gugenheimer, J., Könings, B., Klepsch, M., Schaub, F., Rukzio, E., Seufert, T., & Weber, M. (2014). P.I.A.N.O.: Faster Piano Learning with Interactive Projection. *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 149–158. https://doi.org/10.1145/2669485.2669514

Sweller, J. (1998). *Cognitive Architecture and Instructional Design*.

Wang, Y. (2022). Creative Digital Tools for Teaching Piano. *Revista Música Hodie*, *22*. https://doi.org/10.5216/mh.v22.70961

Wu H.-Y. (2023). *The Effects of Incorporating Tactile Prompts into a Piano Fingering Practice Application on Flow Experience, Cognitive Load, and Learning Emotions for Piano Beginners*.

THE EFFECTS OF TABLE PARTITION TRANSPARENCY ON DINING BEHAVIOR IN PUBLIC SPACES

Pei-Hsuan Wu[a], Meng-Cong Zheng[1a]

[a] *Institute of Innovation Design, National Taipei University of Technology, Taipei, Taiwan, zmcdesign@gmail.com*

ABSTRACT

With the emergence of solo dining culture, customer needs have diverged from those of traditional group diners, and the use of table partitions has generated both positive and negative effects. This study examines changes in dining patterns and proposes a research framework grounded in customers' psychological perceptions and physical seating regulations. A mixed-method approach was adopted, including questionnaire surveys, non-participant observation, and behavioral mapping. The study aims to analyze the effects of table partitions on customer behavior, with the goal of informing future partition designs that reduce dining-related discomfort while addressing solo diners' privacy needs and facilitating smooth conversation among group diners.

The results indicate that customers with different dining types exhibit varying expectations and sensitivities regarding perceived spatial crowding. Diners respond differently in terms of behavior and seating choice when exposed to table partitions with varying levels of transparency. The majority of solo diners prefer seats with opaque partitions and are more tolerant of constrained spatial conditions, showing little inclination to adjust or relocate the partitions. These findings suggest that restaurants may strategically employ different types of table partitions to provide dining environments that are both safe and comfortable for diverse customer groups.

Keywords: *privacy, solo dining, table partitions, perceived crowding, dining satisfaction*

[1] Corresponding author.

1 INTRODUCTION

With shifts in societal demographic structures, a new dining culture—solo dining—has emerged in recent years. Solo dining refers to the practice of individuals eating alone in public dining environments. It has evolved from a “reluctant necessity” into a “self-determined lifestyle choice.” Rather than being perceived as an indicator of loneliness, solo dining is increasingly regarded as a means of enjoying personal time, focusing on food, and engaging in ritualized consumption practices. Previous research has classified solo diners into several categories, including enjoyers, economically motivated diners, social-oriented diners, and relaxation-seeking diners (Koay et al., 2024).

In contrast, shared meals refer to dining experiences in which multiple individuals gather to eat together. According to Baek(2023), shared meals play a significant role in strengthening social bonds, mitigating interpersonal conflict or tension, facilitating intergenerational interaction, and supporting cultural transmission and sustainability.

From the perspective of restaurant operators, however, accommodating solo diners may reduce seating efficiency when operational costs are taken into consideration. Consequently, some restaurants are less inclined to accept single patrons, which may induce psychological pressure among solo diners. The overall dining experience provided by a restaurant—including food quality, service quality, convenience, and spatial atmosphere—significantly influences customers’ perceptions, repurchase intentions, and willingness to recommend the restaurant to others (Richardson et al., 2019).

Due to the boundary effect, individuals tend to seek spatial support elements and prefer locations with clearly defined boundaries (Hsu & Yang, 2021). Solo diners often lack sufficient interpersonal distance or physical partitions from adjacent patrons, and the gaze of nearby diners may reduce their perceived privacy, thereby negatively affecting the dining experience. This observation aligns with the concept of perceived territoriality proposed by Moon et al. (2020), which suggests that territorial perception is shaped not only by physical environmental attributes but also by psychological boundaries. Given the practical limitations of increasing table spacing for solo diners, greater emphasis should be placed on the arrangement of objects at the seating level to establish a more favorable dining environment.

Furthermore, crowding perception is conceptually distinct from environmental density, with the key difference residing in individuals’ psychological perceptions of spatial proximity. As articulated by Stokols (1972), density constitutes a necessary condition for crowding perception but does not, in itself, account for all contributing factors.

Therefore, it is recommended that subsequent research pay attention to (a) behavioral measures pertaining to the augmentation of one's space, (b) subjective reports of restriction and discomfort, (c) observational indexes of tension (e.g., laughter, aggression, reduced eye contact), and (d) physiological indicators of strain.

Andereck and Becker(1993) noted that consumers' expectations regarding environmental density significantly influence their perceived level of spatial crowding. Their study further revealed significant differences between seating configurations and perceived crowding. Compared with side-by-side seating areas, seating types such as high-top seating and multi-person table arrangements were associated with lower levels of perceived crowding among consumers.

Deng(2015) found that seating arrangements were categorized into three types based on social environment theory: general seating, social centrifugal design seating, and social centripetal design seating. Different types of customers in various settings tend to select seating arrangements that best align with their needs.

Lin(2009) investigated the amount of table space required for individuals to dine comfortably and identified the upper limit of personal tabletop territory as 108 cm in width and 63 cm in depth. Although personal tabletop territory varies with table size, it does not continue to increase once this threshold is reached.

The COVID-19 pandemic in 2019 presented both a crisis and an opportunity for the restaurant industry. During the pandemic, restaurants were compelled to implement table partitions as a preventive measure to inhibit virus transmission and maintain safe dining distances, while simultaneously creating physical and social barriers between diners. Even after restrictions were lifted, some establishments continued to use table partitions, as they were found to offer additional functional and operational benefits.

Accordingly, this study seeks to examine whether different transparency designs of table partitions can provide solo diners with a sense of privacy, maintain visual connectivity and conversational openness for group diners, and enhance spatial utilization efficiency.

2 METHODS

The study employed the following research methods for data collection and analysis, including questionnaire survey, non-participant observation, and behavioral mapping. First, a questionnaire survey was conducted to examine participants' prior impressions of restaurant table partitions and their individual sensitivity to the surrounding environment. The experiment primarily utilized non-participant observation and behavioral mapping to investigate diners' behaviors and situational responses under conditions with and without table partitions during the dining process.

This study focused on dining spaces within university campus cafeterias in Taipei City, which had previously been equipped with table partitions. The primary table configuration consisted of four-seat tables measuring 70 cm in depth and 110 cm in width; however, these tables were frequently used by customer groups that did not conform to four-person arrangements, making the site suitable for empirical observation. In addition to serving students for dining, relaxation, social interaction, and reading, the cafeteria was also open to off-campus visitors and local residents.

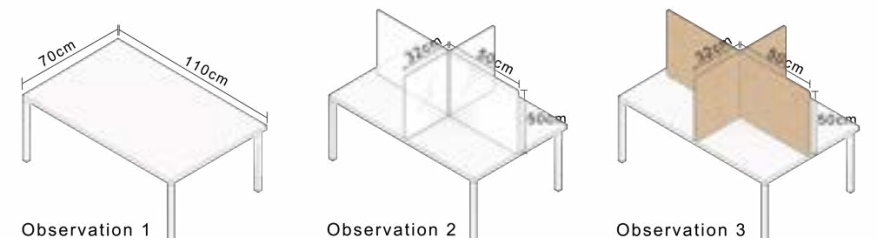


Figure 1 Table partition diagram

A non-participant observation method was employed in the experiment, which was divided into three observation phases: Observation I, Observation II, and Observation III. These were conducted on three separate days under different conditions—without partitions, with transparent partitions, and with opaque partitions, respectively. Observations recorded included group size, frequency of partition adjustments, interaction behaviors, and gaze direction. Based on these observations, inferences were made regarding customers' spatial perceptions, sense of privacy, and conversational convenience, with the aim of understanding current patterns of partition use and evaluating the relationship between observed behaviors and the presence of table partitions.

3 RESULTS

3.1 Questionnaire Survey

When individuals' perceptions of intimacy with others do not align, three types of privacy invasions—physical, visual, and auditory—may occur, resulting in discomfort (Robson, 2008). This study's questionnaire incorporated the Highly Sensitive Person (HSP) Scale developed by Aron et al. (1997), the crowding measurement dimensions proposed by Stokols (1972), and a seven-point Likert scale to quantify participants' subjective experiences.

A total of 104 participants completed the survey, with a gender distribution of 41.3% biologically male and 56.7% biologically female. Participant ages ranged from 11 to 65 years, with the largest age group being 26–30 years (n = 27, 25.9%).

Regarding dining habits, participants reported an average of 5 occasions per week of eating out (dining inside a restaurant), with the most frequently reported frequency being 3 times per week (n = 24, 23.1%). The majority of participants usually dined in pairs (n = 40, 28.4%). Popular types of frequent dining included hot pot (58.6%), Chinese-style snacks (45.1%), and Western-style restaurants (44.2%). The average duration of restaurant visits was 1.4 hours, with the maximum stay not exceeding 3 hours.

Table 1 Reliability Analysis Table for Each Questionnaire Dimension

Item	Scale Dimension	Cronbach's alpha	Number
Part 1	restriction and discomfort	.926	4
Part 2	augmentation of one's space	.85	9
Part 3	Highly Sensitive Person Scale	.888	15

The questionnaire consisted of four sections: (1) restriction and discomfort, (2) augmentation of one's space, (3) Highly Sensitive Person Scale, and (4) open-ended questions. Reliability analysis indicated acceptable internal consistency for all three scale dimensions, with Cronbach's alpha coefficients exceeding 0.85.

3.2 Pearson Correlation analysis

Pearson correlation analyses were conducted between the scores of the three measured dimensions and other variables. The results indicated that restriction and discomfort were significantly negatively correlated with support long-term use of partitions in restaurants [$r(102) = -.332$, $p = .001$]. In the dimension of augmentation of one's space, the tendency to avoid seats near others was significantly positively correlated with support for long-term partition use [$r(102) = .358$, $p = .001$]. Age was significantly negatively correlated with personal sensitivity [$r(102) = -.289$, $p = .003$].

Regarding willingness to support long-term use of partitions, Pearson correlation analysis with other variables revealed that agreement level was significantly negatively correlated with "The presence of partitions restricts my freedom to move my hands or body" [$r(102) = -.369$, $p < .001$], and significantly positively correlated with "I prefer partitions to block the view of others rather than merely separating seats" [$r(102) = .484$, $p < .001$] and "I tend to avoid choosing seats near others" [$r(102) = .358$, $p < .001$]. These findings provide insight into the characteristics of participants who support long-term partition use.

For dining types that involve frequent eating out (dining inside restaurants), Pearson correlation analyses with other variables showed the following patterns:

Hot pot diners were positively correlated with "I do not like people sitting too close to me or talking to me," "I tend to avoid seats near others," and "I do not like too many things happening at once," indicating a preference for spaces that allow for solitary enjoyment.

Light meal diners were positively correlated with "I feel muscle tension or stiffness in such environments," "I sweat or feel anxious when there are too many people around me," "Crowding makes me feel physically uncomfortable or fatigued," "I am easily affected by the emotions of others," "I feel distressed by strong stimuli such as loud noise or chaotic situations," and "Total personal sensitivity score," suggesting a heightened awareness of their surrounding environment.

Chinese snack diners were negatively correlated with "I feel muscle tension or stiffness in such environments," "Crowding makes me feel physically uncomfortable or fatigued," and "Behavioral expansion of personal space and stress total score," indicating less avoidance of close proximity to others and greater suitability for shared dining experiences.

3.3 Qualitative Analysis of Open-Ended Responses

The questionnaire data captured participants' perceptions of the impacts of existing table partitions on their dining experiences. Participants' open-ended responses were analyzed using thematic analysis, resulting in the categorization of user statements. Among the respondents, 14 participants (13.5%) expressed positive perceptions, 52 participants (50%) reported negative

perceptions, and the remaining 38 participants (36.5%) provided neutral responses or indicated no particular opinion.

After consolidating the affective descriptors used in the responses, the most frequently reported negative themes were spatial constraint and discomfort caused by transparency. Neutral themes emphasized the separating function of partitions, noting that while they block the gaze of strangers, they may simultaneously hinder conversation among friends. The most prevalent positive theme was hygiene, followed by perceptions of enhanced privacy and a sense of safety.

3.4 Observation 1

Observation 1 was conducted under existing conditions without table partitions. Pedestrian flow and seating selection changes were recorded at 30-minute intervals during operating hours from 7:30 a.m. to 7:30 p.m. Occupied seats were marked on a floor plan and categorized by group size, including single diners, pairs, groups of three, and groups of four or more. Overlay mapping was then applied to identify areas with the longest duration of solo diner occupancy, the most frequently selected seats, and peak and off-peak traffic periods.

The results showed that among the 153 seats within the observed area, seat utilization was extremely low before 8:00 a.m., at less than 10%. Seat occupancy peaked between 12:00 p.m. and 12:30 p.m., reaching 50%. The highest number of solo diners was observed between 11:30 a.m. and 12:30 p.m., after which pedestrian flow declined noticeably around 1:30 p.m.

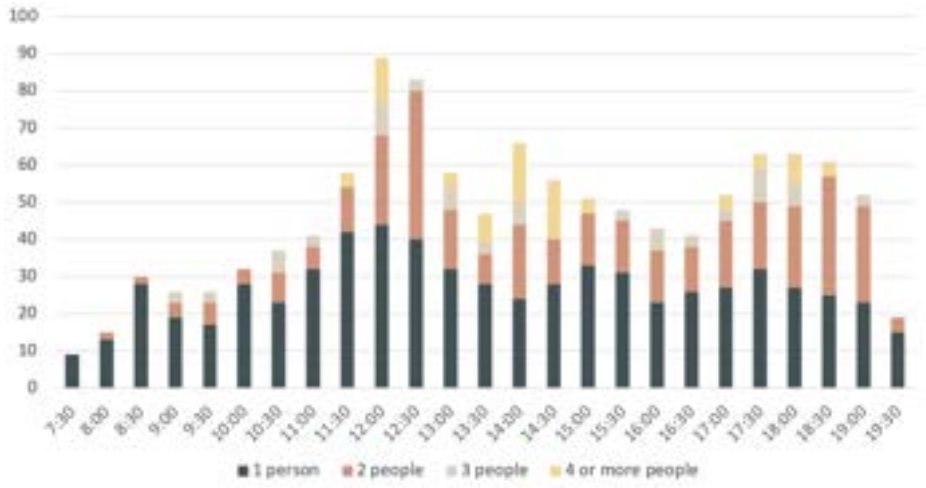


Figure 2 Cumulative bar chart of the number of diners without partitions

For seats selected by solo diners, occupied seats were recorded every 30 minutes. In this study, the Green Courtyard seating area was divided into areas, each consisting of three four-person tables, and usage frequencies were aggregated accordingly. The areas were labeled A through K, as shown in Figure 4.2. The statistical results indicated that Area D, located in front of the convenience store, had the highest frequency of solo diner occupancy (73 instances), followed by Area E, situated near the entrance and adjacent to a column (60 instances).

Subsequent Observation 2 and Observation 3 were designed based on these findings. On two different days, observations were conducted between 7:30 a.m. and 1:30 p.m., covering transitions from off-peak to peak and back to off-peak periods, in order to examine variations in pedestrian flow and partition use. On each day, two high-traffic areas—Area D in front of the convenience store and Area E near the entrance—were equipped with either fully transparent or opaque partitions to increase the number of observable samples.

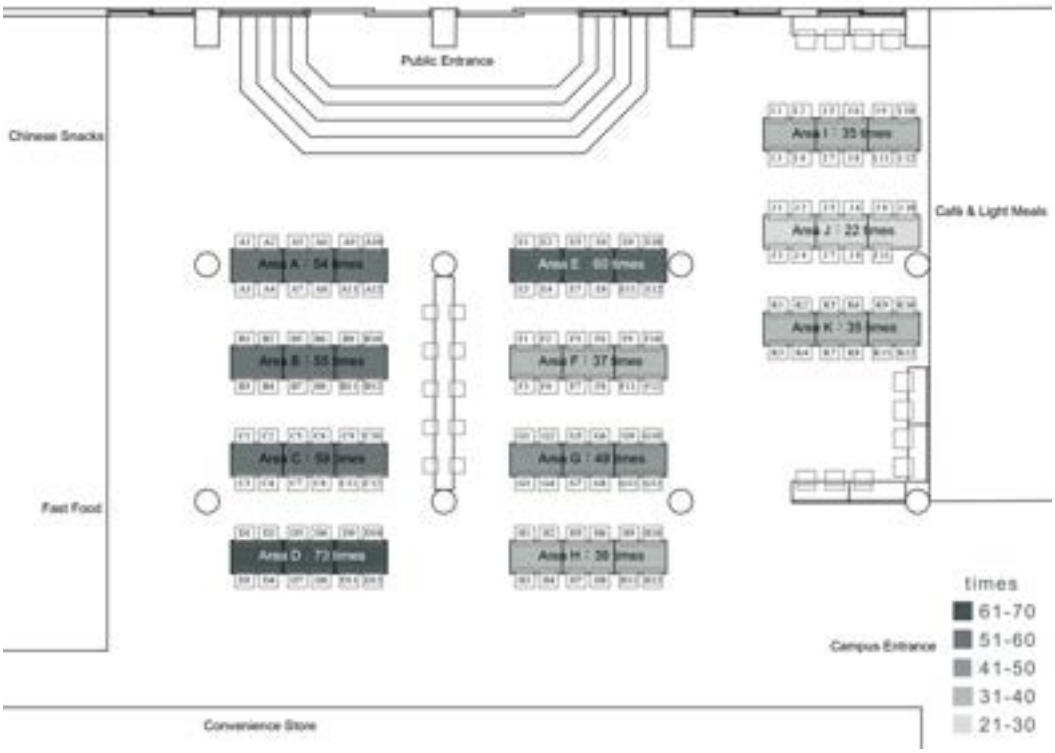


Figure 3 Distribution of seating duration for solo diners without partitions

3.5 Observation 2

Observation 2 was conducted with transparent table partitions installed in two areas: Area D (12 seats) and Area E (12 seats). During operating hours from 7:30 a.m. to 1:30 p.m., pedestrian flow, seating selection, seat-occupying behaviors, and individual behavioral changes within these two areas were recorded at 10-minute intervals. Occupied seats and observed behaviors were marked on a floor plan, and overlay mapping was applied to calculate seating selection proportions as well as to analyze seat occupation and individual behavioral patterns.

Statistics on customer types in Areas D and E indicate that solo diners accounted for the largest proportion, comprising 57% (70 instances) in Area D and 64% (56 instances) in Area E. Other shared-meal diners were randomly distributed across the seating areas and were not limited to areas primarily used by solo diners.

3.6 Observation 3

Observation 3 was conducted with opaque table partitions installed in two areas: Area D (12 seats) and Area E (12 seats). During operating hours from 7:30 a.m. to 1:30 p.m., pedestrian flow, seating selection, seat-occupying behaviors, and individual behavioral changes within these two

areas were recorded at 10-minute intervals.The distribution of customer types in Areas D and E showed that solo diners accounted for the highest proportions, at 85% (89 instances) and 75% (60 instances), respectively. These proportions were markedly higher than those observed when transparent partitions were installed. In contrast, no customers dining in groups of three or four or more (0 instances) selected seating in either of these two areas.

Table 2 Statistics of Diners by Group Size in Areas D and E under Different Conditions

Number of people dining together	Transparent Partition		Opaque Partition	
	Area D	Area E	Area D	Area E
1 person	70	56	89	60
2 people	7	28	16	20
3 people	27	0	0	0
4 or more people	20	4	0	0

Because some groups had zero counts in the ''three-person'' and ''four-or-more-person'' categories, dining group size was recoded as ''≤ 2 persons'' and ''≥ 3 persons'' to avoid low expected frequencies and to focus on group-size effects. Fisher's exact test was conducted separately for each area to examine the association between partition transparency (transparent vs. opaque) and the proportion of larger groups (≥ 3 persons). In addition, under the transparent-partition condition, proportional differences between Areas D and E were compared.

In Area D, the proportion of ≥ 3-person dining under transparent partitions was 37.9% (47/124), compared with 0% (0/105) under opaque partitions; Fisher's exact test indicated a significant difference ($p < .001$). In Area E, the proportion of ≥ 3-person dining under transparent partitions was 4.5% (4/88), compared with 0% (0/80) under opaque partitions; this difference did not reach significance ($p = .122$). Furthermore, under the transparent-partition condition, the proportion of ≥ 3-person dining was significantly higher in Area D than in Area E (37.9% vs. 4.5%, $p < .001$), indicating that larger groups predominantly appeared in transparent-partition settings, with this pattern being more pronounced in Area D.

In Area D and E, customer behaviors were recorded at 10-minute intervals and aggregated under conditions with opaque partitions installed. To examine whether the distribution of customer behavior categories was consistent across the four conditions (Area D with transparent partitions, Area D with opaque partitions, Area E with transparent partitions, and Area E with opaque partitions), Pearson's chi-square test was employed. The results indicated a significant association between experimental condition and the distribution of customer behaviors, $\chi^2(15, N = 426) = 38.20$, $p = .001$, demonstrating that customer behavior distributions differed across conditions. Further comparisons revealed a significant difference in behavior distributions between Area D with transparent partitions and Area D with opaque partitions, $\chi^2(5, N = 254) = 13.62$, $p = .018$. This finding indicates that, within the Area D context, partition transparency is more strongly associated with variations in customer behavior distributions.

A contingency table was constructed with behavior categories (7 types) and contextual conditions (4 types: transparent/opaque × Areas D/E), and Pearson's chi-square test of independence was conducted. The results indicated that the distribution of customer behaviors differed significantly across the four conditions, $\chi^2(18, N = 427) = 40.17$, $p = .00198$, with Cramér's $V = .177$. Given the presence of small expected frequencies, a fixed-margin Monte Carlo simulation (50,000 iterations) yielded consistent results ($p = .00128$), demonstrating the robustness of the observed differences.

Further examination using adjusted standardized residuals revealed that "Dining-related" behaviors were significantly higher than expected in the Opaque–Area E condition and significantly lower than expected in the Transparent–Area D condition. In addition, the Transparent–Area D condition exhibited higher-than-expected frequencies of "Interaction among co-diners," "One-way interaction," and "Temporarily put" behaviors. Overall, the contextual conditions not only produced statistically significant differences in behavioral proportions with a modest effect size ($V = .177$), but these differences were also concentrated in the relative increases and decreases of "Dining-related" and "Interaction / Temporarily put" behaviors under specific conditions. These findings suggest that partition visibility and area-specific characteristics may jointly influence customers' patterns of stay and interaction within the dining environment.

Table 3 Frequency Table of Customer Behaviors across Different Conditions

Behavior	Transparent Partition		Opaque Partition	
	Transparent – Area D (n=142)	**Transparent – Area E (n=91)**	**Opaque – Area D (n=112)**	**Opaque – Area E (n=82)**
Dining-related	33 (23.2%) ↓ *	23 (25.3%)	36 (32.1%)	38 (46.3%) ↑ *
Interaction among co-diners	28 (19.7%) ↑ *	14 (15.4%)	12 (10.7%)	8 (9.8%)
One-way interaction	5 (3.5%) ↑ *	1 (1.1%)	1 (0.9%)	0 (0.0%)
Personal behavior	54 (38.0%)	46 (50.5%)	50 (44.6%)	32 (39.0%)
Looking down	1 (0.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Temporarily put	19 (13.4%) ↑ *	3 (3.3%)	7 (6.2%)	2 (2.4%)
Sleeping	2 (1.4%)	4 (4.4%)	6 (5.4%)	2 (2.4%)

Pearson $\chi^2(18, N = 427) = 40.17$, $p = .00198$；Monte Carlo（50,000 times）$p = .00128$；Cramer's $V = .177$。

4 CONCLUSION

Based on the results of the questionnaire survey, this study analyzed participants' psychological expectations through three dimensions: subjective restriction and discomfort, augmentation of one's space, and personal sensitivity. The findings indicated that customers with different dining

types—including hot pot, light meals, and Chinese snack diners—exhibited varying expectations regarding spatial crowding and environmental sensitivity.

In addition, using thematic analysis, participants' positive and negative perceptions of table partitions were categorized. Negative perceptions were primarily associated with spatial constraint and discomfort caused by transparency, while positive perceptions emphasized hygiene, privacy, and a sense of safety.

Following the non-participant observation experiment, the relationship between table partitions and customer dining behavior was analyzed. The results showed that diners exhibited different behaviors and seating choices depending on the transparency of the table partitions, with behavioral distributions demonstrating significant differences. During the observations, the majority of solo diners preferred seats with opaque partitions and were willing to accept constrained space without relocating the partitions. Based on these findings, it is recommended that restaurants strategically employ different types of table partitions to provide safe and comfortable dining environments tailored to diverse customer needs.

REFERENCES

Andereck, K. L., & Becker, R. H. (1993). Perceptions of carry - over crowding in recreation environments. Leisure Sciences, 15(1), 25 – 35.

Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. Journal of Personality and Social Psychology, 73(2), 345–368.

Baek, S.-J. (2023). The implications and social functions of shared meals in Japanese society. Sig Gonggan Yeon-Gu, 18(2), 1–17.

Denga, P. C., & Zheng, M. C. (2014). The Different Approach Toward Social Centrifugal and Social Centripetal Designed Public Seating in International Airports. [Master's thesis, National Taipei University of Technology]. Taiwan Digital Theses and Dissertations System.

Hsu, L.-Q., & Yang, G.-X. (2021). Environmental psychology: Environment, perception, and behavior. Wunan Books.

Koay, K. Y., Ong ,Derek Lai Teik, Idris ,Izian, Kamarulzaman ,Nurfarah Sabrina Binti, & and Cheah, C. W. (2024). The rise of solo dining: Prediction and consumer profiling. Journal of Foodservice Business Research, 1–20.

Lin, U.-T. (2009). A Fundamental Study of Personal Tabletop Territory. [Master's thesis, National Taipei University of Technology]. Taiwan Digital Theses and Dissertations System.

Moon, S., Bonn, M. A., & Cho, M. (2020). How can the solo dining experience be enhanced? Focusing on perceived territoriality. International Journal of Hospitality Management, 88, 102506.

Richardson, S., Lefrid, M., Jahani, S., Munyon, M. D., & Rasoolimanesh, S. M. (2019). Effect of dining experience on future intention in quick service restaurants. British Food Journal, 121(11), 2620–2636.

Robson, S. K. A. (2008). Scenes from a restaurant: Privacy regulation in stressful situations. Journal of Environmental Psychology, 28(4), 373–378.

Stokols, D. (1972). On the distinction between density and crowding: Some implications for future research. Psychological Review, 79(3), 275–277.

THE IMPACT OF VISUAL VARIABLES IN FLOOD HAZARD MAP ON THE PERCEPTION OF EURO-ASIAN CULTURAL GROUPS

Ai Ruo Chang[a], Meng-Cong Zheng[1b]

[a] *Department of Industrial Design, National Taipei University of Technology, Taiwan, t113588009@ntut.org.tw*
[b] *Department of Industrial Design, National Taipei University of Technology, Taiwan, zmcdesign@gmail.com*

ABSTRACT

As flood events become more frequent, Flood Hazard Map (FHM) play a key role in transnational disaster management. Yet prior research shows that difference in cultural background affect how users interpret disaster information, particularly comprehension, visual aesthetics, and risk perception, the absence of consistent design guidelines for flood data leads to misinterpretations in cross-border communication. Therefore, this study compares European and Asian users to identify design directions for cross-cultural flood risk communication.

This study developed 6 Flood hazard map samples using two visual variables (color scheme × number of color) and conducted an online survey with 240 participants from Taiwan, Japan, the United Kingdom, and the Czech Republic, using validated scales and preference rankings across three dimensions. Results revealed cross-cultural differences in Flood hazard map-use behavior: Europeans more frequently used Flood hazard map for frequently visited places and travel destinations, indicating stronger cross-contextual risk-checking tendencies, whereas Asians focused more on residential-area risk identification. Regarding visual variables performance, the 4-levels blue–red scheme consistently ranked among the most preferred options across both regions. However, objective evaluations diverged: European participants rated option Ca (Yellow-Red + 4-levels) highest on most measures, except DAS-Adequacy of Information, while Asians showed the opposite pattern, rating Ca lowest in usability outcomes.

Based on the results, this research integrates the performance of visual variable across comprehension, visual aesthetics, and risk perception. It further examines factors shaping the relationship between scale and ranking, and the results across the three perceptual dimensions.

[1] Corresponding author. 1, Sec. 3, Zhongxiao E. Rd.,Taipei 10608 Taiwan, zmcdesign@gmail.com

Finally, the findings are translated into design recommendations, underscoring the importance of a professional yet simplified strategy to prevent information overload and enhance risk communication.

Keywords: *Visualized flood information, cross-cultural design, individual perception, Mixed-quantitative analysis*

1 INTRODUCTION

Among various natural disasters, floods cause the most severe damage to both the natural environments and socio-economic systems. Compared with disasters such as earthquakes and droughts, floods have broader spatial impacts and occur with greater frequency. Statistics indicate that between 1990 and 2022, over 4,713 flood events occurred worldwide, affecting approximately 3.2 billion people and resulting in economic losses exceeding 1.3 trillion USD (Liu et al., 2024). In light of this, the United Nations Office for Disaster Risk Reduction (UNDRR), recommended in the *Sendai Framework for Disaster Risk Reduction 2015–2030* that countries should utilize Geographic Information Systems (GIS) to promote disaster risk assessment (UNDRR, 2015). Consequently, the "Flood Hazard Map" are a medium that uses visual elements to present flood data, and are widely used in decision-making, risk understanding, have become an essential tool for national risk communication and disaster management globally (GWP & WMO, 2013).

High-risk areas influenced by climate change are primarily concentrated in densely populated regions such as East Asia, South Asia, and Central–Western Europe (Duan et al., 2022). Accordingly, both Europe and Asia have developed their Flood hazard map planning strategies in response to these conditions, exhibiting significant regional differences. In Asia, Flood hazard map are largely government-led with a primary focus on emergency early warning systems, such as Taiwan's Disaster Potential Maps and Japan's unified visualize planning for flood hazards. Conversely, European countries, driven by the EU Floods Directive (2007/60/EC), have transitioned from a "security-oriented" approach to "risk-based" thinking, focusing on flood risk management and enhancing overall flood awareness across Europe (EU, 2007; Müller, 2013).

However, relevant research indicates that due to cultural differences and geographical environments, users across countries differ in efficiency and cognitive logic when searching for visualized information on maps. Such cognitive disparities not only affect Flood hazard map's comprehension but also influence judgments of visual aesthetics and perceived risk severity, ultimately shaping disaster response decisions and behavioral intentions (Razak et al., 2025; Stieb et al., 2019). Therefore, to accurately measure how cultural factors intervene in the effectiveness of risk communication, this study identifies comprehension, visual aesthetics, and risk perception as the three dimensions for evaluating cross-cultural perception levels:

1. Comprehension: The complexity of map information directly affects cognitive load during reading. Poor design can lead to excessive cognitive demand. Therefore, cartographers often draw on cognitive load theory to examine whether visual variables (e.g., color variations, presence of symbols) effectively reduce cognitive

burden, enhance information processing efficiency, and improve decision accuracy (Al-Samarraie et al., 2019; Bunch & Lloyd, 2006).

2. Visual Aesthetics: Visual aesthetics constitute a central component of the map-use experience. Beyond improving data recognition, aesthetics can foster users' emotional engagement with disaster contexts (Nestel, 2025). Users' aesthetic perceptions of maps primarily derive from visual clarity and consistency (Imus & Loftin, 2012).
3. Risk Perception: The perceived severity of disasters depends partly on the meaning conveyed by the event and partly on the selection of visual elements. For instance, in Western cultures, the use of psychologically salient colors such as red—associated with danger and warning—and their tonal adjustments can effectively heighten viewers' awareness of risk levels (Slovic et al., 1984; Welhausen, 2015).

Current research primarily focuses on data accuracy, while comparatively limited attention has been given to how visual variables influence perception across cultural groups. Therefore, this study aims to investigate differences in how visual variables affect individual perception across European and Asian cultural contexts, with the goal of deriving generalizable design recommendations to enhance the effectiveness of international risk communication.

2 METHODS

2.1 Participants

The survey targeted representative countries in Asia and Europe. Asian participants include Taiwan (Mandarin-speaking country) and Japan (non-Mandarin-speaking country), while European participants consist of the United Kingdom (English-speaking country) and the Czech Republic (non-English-speaking country). The recruitment criteria for this survey required participants to be aged 18 years or older and to have no color vision deficiencies. A total of 60 participants were recruited in each country, resulting in 240 participants overall.

2.2 Samples Design

Based on flood hazard map guidelines in Europe and Asia, experimental variable were designed focusing on two core visual variables: "flood hazard color schemes" and "number of color levels." The color schemes were selected based on common schemes across various Asian and European countries, resulting in three distinct binary color schemes. The number of color level was set at four and 6-levels to examine the effects of color quantity on risk discrimination and cognitive load. Consequently, a total of 6 samples were generated to investigate differences in individual perception across cultural backgrounds. (Figure 1)

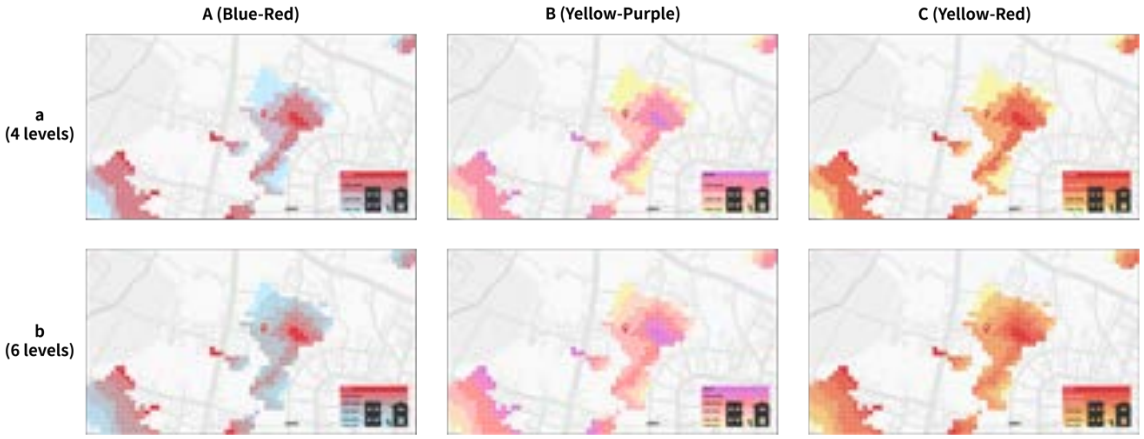


Figure 1. Survey Smaples with a 2 × 3 variable design

2.3 Questionnaire Design

An online structured questionnaire was employed to assess participants' comprehension, visual aesthetics, and risk perception of the visual variables. The questionnaire consisted of three stages: (1) background information of participant, followed by measures of three perceptual dimensions using (2) rating scales and (3) preference ranking.

2.3.1 Background Information

The survey collected participants' information, included gender, age group, nationality, and the presence of any color vision deficiency. In addition, a 7-point Likert scale was used to assess participants' experience in reading flood hazard maps across contexts, including checking risks near residence, frequently visited areas, during heavy rain forecasts, and for travel destinations.

2.3.2 Scales

This phase focuses on the three perceptual dimensions, employing one to two scales for each. For the comprehension dimension, the "understand" and "dataread" sub-scales of the PREVis (Perceived Readability Evaluation for Visualizations)—comprising a total of 6 items—were used to evaluate the ease with which users read visualized information on the maps (Cabouat et al., 2025). Additionally, the System Usability Scale (SUS) was utilized, featuring 10 alternating positive and negative items on a 5-point Likert scale, to assess the overall perceived usability of the maps (Lewis, 2018).

In the visual aesthetics dimension, the Visual Aesthetics of Websites Inventory Short version (VisAWI-S) was adopted, explores the impact of visual variables on aesthetic appeal through four facets: simplicity, diversity, colorfulness, and craftsmanship (Moshagena & Thielsch, 2015).

For the risk perception, the "usability of information" and "adequacy of information" facets of the Decision Attitude Scale (DAS), combined with the "Perceived Severity" dimension of the Risk Behavior Diagnosis scale (RBDs), were utilized to assess participants' subjective risk perception within flood scenarios (Sainfort & Booske, 2000; Witte, 1996).

2.3.3 Ranking

After completing the scale evaluations for each dimension, all experimental samples were displayed simultaneously. Participants were requested to rank their personal preferences for comprehension, visual aesthetics, and risk perception based on the all combinations of visual variables. The collected ranking data were used to analyze whether European and Asian participants differed in their preference ordering after viewing identical flood hazard maps.

3 RESULT

In addition to collecting demographic information and Flood hazard map-use experience, this survey employed both rating scales and ranking tasks to examine European and Asian users' objective evaluations and subjective preferences for samples, with data analyzed using SPSS software.

3.1 Background Information

A total of 240 valid questionnaires were collected. Of these, 120 respondents were from Europe (60 from the Czech Republic and 60 from the United Kingdom) and 120 from Asia (60 from Japan and 60 from Taiwan). The gender distribution was relatively balanced (55.8% male, 44.2% female), most participants were aged 18–30 (35.4%) and 31–40 (23.3%). All participants reported no color vision deficiency, ensuring the validity of the experiments.

Based on the survey of flood hazard map consultation experience, European participants (M=3.53, SD=1.95) scored significantly higher than Asian participants (M=2.90, SD=1.82, p=.011) when checking flood risks for frequently visited areas. Similarly, in the context of checking risks for travel destinations, the European group (M=3.38, SD=1.85) also scored significantly higher than the Asian group (M=2.40, SD=1.52, p<.001), indicating that European users demonstrate stronger cross-contextual risk-checking behaviors when using flood hazard maps. One-way ANOVA further highlighted significant country-level variations: in the "frequently visited" context (F=3.63, p=.014), with the United Kingdom showing a significantly higher frequency than Taiwan. In the "travel destination" context, country differences were also significant (F=7.356, p<.001), with UK participants again scoring significantly higher than those from Taiwan and Japan.

3.2 Scale Results

An two-way and three-way ANOVA was conducted to examine the interactions between two regions user and experimental samples, as well as the effects of two visual variables within the scale result. Based on the perceptual dimensions: comprehension, visual aesthetics, and risk perception, explains how visual variables influence European and Asian users.

3.2.1 Comprehension

The PREVis results for the two subscales: understand and dataread revealed that the European group showed high agreement for sample Ca (Yellow-Red + 4-levels), with high ratings in both understand (M=5.32, SD=0.933) and dataread (M=4.98, SD=1.137) subscales, suggesting that this configuration provides greater visual stability. In contrast, the Asian group gave Ca the lowest rating for understand (M=4.60, SD=1.358), with the most preferred samples for understand being

Aa (Blue-Red + 4-levels; *M*=5.42, *SD*=1.09), and the highest-rated samples for dataread being Ba (Yellow-Purple + 4-levels ; *M*=5.08, *SD*=1.19).

Regarding the SUS result, the usability grades for all samples in both regions fell within grade D or F, indicating significant room for optimization in current Flood hazard map despite the inclusion of visualized information. In the European group, Aa and Ca received the highest ratings (*M*=64.38, SD_{Aa}=15.43, SD_{Ca}=14.75), whereas the Asian group gave the lowest rating to the sample Ca (*M*=52.00, *SD*=15.65); instead, Ba achieved the highest score in Asia (*M*=64.75, *SD*=9.42). The three-way ANOVA results (Table 1) further confirmed a significant main effect of the number of color levels (*F*=3.90, *p*=.049) and a significant interaction between color scheme and number of color with a small effect size (*F*=4.96, *p*=.008, η_p^2=.042), indicating that usability evaluations of Flood hazard map in the comprehension-related dimension were primarily influenced by combinations of visual variables, while the overall trends remained largely consistent between European and Asian groups with no significant differences (*p*=.167).

Table 1. Three-way ANOVA of visual variables & cultural groups in the SUS (N=240)

	SS	***df***	***MS***	***F***	***p***	η_p^2
A. Color Schemes	130.20	2	65.10	0.25	.774	.002
B. Number of Color	990.23	1	990.23	3.90*	.049	.017
C. Cultural Groups	25.02	1	25.02	0.09	.754	.000
A*B	2520.62	2	1260.31	4.96**	.008	.042
A*C	851.45	2	425.72	1.67	.189	.015
B*C	446.90	1	446.90	1.76	.186	.008
A*B*C	915.83	2	457.91	1.80	.167	.016
Error	57834.68	228	253.66			
Total	911293.75	240				
Corrected Total	63714.97	239				

*p<.05, **p<.01, ***p<.001.

3.2.2 Visual Aesthetics

The mean VisAWI-S scores indicated that the European group gave the highest aesthetic rating to sample Ca (Yellow-Red + 4-levels; *M*=4.99, *SD*=1.28), and also evaluated all 4-levels samples (Aa, Ba, Ca) more positively. In contrast, the Asian group showed the highest preference for sample Cb (Yellow-Red + 6-levels; *M*=5.14, *SD*=0.94). When comparing the mean scores across individual items, both European and Asian participants reported the highest agreement for “Craftsmanship”, with mean scores of 4.81 and 4.83, respectively. In contrast, "Diversity" received the lowest ratings, with a mean of 4.63 for European group and 4.46 for Asian group, suggesting that while the map visuals effectively conveyed professionalism and credibility, there remains room to enhance visual variation in design.

3.2.3 Risk Perception

The DAS results showed that, in terms of “Usability of Information”, the European group gave the highest rating to sample Ca (Yellow-Red + 4-levels; *M*=3.83, *SD*=0.49), whereas the Asian group rated Aa highest (Blue-Red + 4-levels; *M*=3.85, *SD*=0.72), this pattern is consistent with the PREVis’s understand subscales findings. Regarding “Adequacy of Information”, scores for the 6-levels color samples were generally higher than those for the 4-levels samples , such as Europe highest rating sample Cb (Yellow-Red + 6-levels; *M*=2.75, *SD*=0.83); Asia highest rating sample Ab (Blue-Red + 6-levels; *M*=2.68, *SD*=0.93), suggesting that finer color stratification enhances users’ perceived adequacy of information.

According to the RBDs scale results, although there was no significant difference in the level of risk perception across samples between European and Asian participants (*p*=.296), mean scores indicated that both the European (*M*=3.88, *SD*=0.63) and Asian (*M*=3.97, *SD*=0.71) groups exhibited higher risk perception for sample Ca.

3.3 Preference Ranking

Based on the Friedman test of participants’ subjective preference rankings (Table 2), the results showed that except for the Asian group’s rankings on visual aesthetics which were not significant (*p*=.704), both European and Asian users demonstrated significant differences across the remaining three perceptual dimensions (*p*<.001).

For comprehension, the Blue-Red sample Aa and Ab occupied the first and second ranks. However, the least comprehensible sample differed by region: for the European group, Cb (Yellow-Red + 6-levels; *M*=2.94, *SD*=1.94) ranked lowest, whereas for the Asian group, Bb (Yellow-Purple + 6-levels) was rated as the least comprehensible (*M*=2.87, *SD*=1.46). In the visual aesthetics dimension, significant differences were observed only in the European group, with Aa ranked highest (Blue-Red + 4-levels; *M*=4.22, *SD*=1.89) and Cb ranked lowest (*M*=2.93, *SD*=1.82).

For risk perception, Aa was the highest-ranked sample for the European group (*M*=4.09, *SD*=1.87), while Ca ranked highest in the Asian group (*M*=4.05, *SD*=1.54). Notably, both groups ranked the Yellow-Purple sample (Ba, Bb) lowest, suggesting that this color scheme is less effective at triggering disaster awareness across cultures. Overall, the results indicate that users generally prefer clear and professional visual configurations, and that the Blue-Red color scheme combined and simplified color gradations demonstrates strong cross-cultural stability.

Table 2. Friedman test of users' rankings across the three perceptual dimensions

	Comprehension		**Visual Aesthetics**		**Risk Perception**	
	Europe (*n*=120)	**Asia (*n*=120)**	**Europe (*n*=120)**	**Asia (*n*=120)**	**Europe (*n*=120)**	**Asia (*n*=120)**
Aa(Blue-Red+4 Levels) , *M(SD)*	4.46(1.74)	4.26(1.78)	4.22(1.89)	3.55(1.98)	4.09(1.87)	3.98(1.80)
Ab(Blue-Red+6 Levels) , *M(SD)*	4.07(1.52)	3.80(1.68)	3.88(1.51)	3.35(1.75)	3.72(1.71)	3.79(1.62)
Ba(Yellow-Purple +4 Levels) , *M(SD)*	3.21(1.39)	2.98(1.55)	3.17(1.38)	3.57(1.53)	3.12(1.42)	2.48(1.46)
Bb(Yellow-Purple +6 Levels) , *M(SD)*	2.96(1.37)	2.87(1.46)	3.13(1.55)	3.70(1.59)	3.09(1.43)	2.81(1.46)
Ca(Yellow-Red+4 Levels) , *M(SD)*	3.37(1.63)	3.72(1.61)	3.67(1.69)	3.39(1.55)	3.53(1.59)	4.05(1.54)
Cb(Yellow-Red +6 Levels) , *M(SD)*	2.94(1.94)	3.37(1.75)	2.93(1.82)	3.43(1.80)	3.45(1.95)	3.88(1.94)
df	5	5	5	5	5	5
x^2	66.77***	47.56***	43.22***	2.97	24.25***	78.17***
p	.000	.000	.000	.704	.000	.000

*p<.05, **p<.01, ***p<.001.

4 DISCUSSION

According to the research findings, the effects of visual variables influence comprehension, visual aesthetics, and risk perception are not entirely consistent across cross-cultural contexts. For the comprehension, the 6-levels color led to a decline in performance for both European and Asian participants, indicating that simplifying visual hierarchies provides a cross-cultural benefit. In contrast, clearer cultural differences emerged in visual aesthetics: while the yellow–red scheme received relatively high scale ratings in the European sample, Asian participants rated it lower and instead showed more positive judgments toward the blue–red and yellow–purple schemes, reflecting how cultural differences shape user responses to color configurations. Regarding risk perception, 4-levels color samples outperformed 6-levels ones in scale scores, indicating that simplified visual structures with clear contrast facilitate the identification of hazard information. Overall, the 4-levels color consistently outperformed the 6-levels across all dimensions, while blue–red scheme demonstrates the highest subjective cross-cultural stability.

To clarify deeper factors underlying cross-cultural evaluations, Spearman correlation analyses revealed that across the three perceptual dimensions, subjective preference rankings generally did not show significant correlations with objective scale scores. This suggests that users may rely on two distinct evaluative logics when interpreting Flood hazard map: one grounded in rational endorsement based on information-reading efficiency, and the other driven by subjective preference shaped by immediate visual impressions and culturally embedded intuitions. This

dual-track pattern implies that relying on only one type of evaluation may be insufficient to comprehensively capture cross-cultural map-use experiences.

Pearson correlation analyses further confirmed significant positive relationships among comprehension, visual aesthetics, and risk perception (Table 3). This demonstrates that regardless of cultural background, aesthetic elements in flood hazard maps can enhance perceived readability and trust, thereby positively strengthening group-level risk cognition and decision intention. However, DAS "Adequacy of Information" exhibited significant negative correlations with two sub-scales of PREVis, VisAWI-S, DAS "Usability of Information", and RBDs scale (r=−.181 to −.366, p<.01), indicating that information overload may instead undermine comprehension and aesthetic appraisal, ultimately weakening overall risk communication effectiveness. The results highlight that achieving an appropriate balance between informational completeness and cognitive load constitutes a challenge in flood hazard map design.

Table 3. Pearson correlation analysis among the scales of each perceptual dimension (N=240)

	PREVis-Understand	PREVis-Dataread	SUS	VisAWI-S	DAS-Usability	DAS-Adequacy
PREVis-Understand	-					
PREVis-Dataread	.826**	-				
SUS	.649**	.624**	-			
VisAWI-S	.542**	.500**	.482**	-		
DAS-Usability	.625**	.584**	.647**	.628**	-	
DAS-Adequacy	-.207**	-.181**	-.054	-.366**	-.227**	-
RBDs scale	.385**	.363**	.312**	.505**	.498**	-.493**

*p<.05, **p<.01, ***p<.001.

5 CONCLUSION

This study investigated cross-cultural perceptual differences between European and Asian users under the influence of visual variables in flood hazard maps, and identified the most influential combinations of visual variables based on both subjective preferences and objective evaluations. The findings further verified a decoupling phenomenon between "rational endorsement" and "intuition-driven judgment" in users' perception of Flood hazard map, indicating that Flood hazard map planners should not focus solely on functional usability, but also incorporate the guidance of visual intuitions in map design.

In addition, the study demonstrates positive interactions among comprehension, visual aesthetics, and risk perception, highlighting the value of aesthetic optimization in strengthening decision intentions. Nevertheless, given the detrimental effects of information overload on perceptual performance, Flood hazard map design should prioritize a progressive information disclosure strategy (from simple to complex), and establish clear hierarchies between core overviews and deeper details. This approach can satisfy user's need for sufficient information while preventing communicative effectiveness from being compromised by redundant content.

Ultimately, such improvements can enhance public interpretation efficiency and contribute to building a consensus-based visual language for flood risk communication across regions.

ACKNOWLEDGMENTS

This research expresses sincere gratitude to Professor Naoyuki Oi of Kyushu University, as well as Professor Vladimír Kovařík and Professor Pavel Krutil of Tomas Bata University in Zlín for their assistance during the research, which enabled the successful completion of questionnaire survey.

REFERENCES

Al-Samarraie, H., Eldenfria, A., Price, M. L., Zaqout, F., & Fauzy, W. M. (2019). Effects of map design characteristics on users' search performance and cognitive load: An empirical study. *The Electronic Library*, *37*(4), 667–679. (world). https://doi.org/10.1108/EL-10-2018-0202

Bunch, R. L., & Lloyd, R. E. (2006). The Cognitive Load of Geographic Information. *The Professional Geographer*, *58*(Issue 2), 209–220.

Cabouat, A.-F., He, T., Isenberg, P., & Isenberg, T. (2025). PREVis: Perceived Readability Evaluation for Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *31*(1), 1083–1093. https://doi.org/10.1109/TVCG.2024.3456318

Duan, Y., Xiong, J., Cheng, W., Li, Y., Wang, N., Shen, G., & Yang, J. (2022). Increasing Global Flood Risk in 2005–2020 from a Multi-Scale Perspective. *Remote Sensing*, *14*(21), 5551. https://doi.org/10.3390/rs14215551

Endo, R., & Nakano, T. (2019). Improvement suggestions for problems of hazard map from the viewpoint of color scheme. *Proceedings of the ICA*, *2*, 1–7. https://doi.org/10.5194/ica-proc-2-25-2019

EU. (2007). *DIRECTIVE 2007/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2007 on the assessment and management of flood risks*. http ://data.europa.eu/eli/dir/2007/60/oj

GWP & WMO. (2013). *Flood Mapping*. World Meteorological Organization (WMO). https://library.wmo.int/idurl/4/37083

Imus, D., & Loftin, P. (2012). The Beauty of Clear Communication. *Cartographic Perspectives*, (73), Article 73. https://doi.org/10.14714/CP73.591

Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Liu, Q., Du, M., Wang, Y., Deng, J., Yan, W., Qin, C., Liu, M., & Liu, J. (2024). Global, regional and national trends and impacts of natural floods, 1990–2022. *Bulletin of the World Health Organization*, *102*(6), 410–420. https://doi.org/10.2471/BLT.23.290243

Moshagena, M., & Thielsch, M. T. (2015). *Visual Aesthetics of Websites Inventory* [Dataset]. https://doi.org/10.1037/t25444-000

Müller, U. (2013). Implementation of the flood risk management directive in selected European countries. *International Journal of Disaster Risk Science*, *4*(3), 115–125. https://doi.org/10.1007/s13753-013-0013-y

Nestel, C. (2025). Theorizing 'Really Good Maps': Cartography and Aesthetic Philosophy. *The Cartographic Journal*, *0*(0), 1–16. https://doi.org/10.1080/00087041.2024.2446051

Razak, S. F. A., Yogarayan, S., Bukar, U. A., & Sayeed, M. S. (2025). Managing flash flood crises with cultural perspectives: A user-centric feature identification study. *PLOS ONE*, *20*(2), e0318996. https://doi.org/10.1371/journal.pone.0318996

S. Seipel & N. J. Lim. (2017, February 15). *Color map design for visualization in flood risk assessment*. https://www.tandfonline.com/doi/full/10.1080/13658816.2017.1349318#abstract

Sainfort, F., & Booske, B. C. (2000). Measuring Post-decision Satisfaction. *Medical Decision Making*, *20*(1), 51–61. https://doi.org/10.1177/0272989X0002000107

Slovic, P., Lichtenstein, S., & Fischhoff, B. (1984). Modeling the Societal Impact of Fatal Accidents. *Management Science*, *30*(4), 464–474. https://doi.org/10.1287/mnsc.30.4.464

Stieb, D. M., Huang, A., Hocking, R., Crouse, D. L., Osornio-Vargas, A. R., & Villeneuve, P. J. (2019). Using maps to communicate environmental exposures and health risks: Review and best-practice recommendations. *Environmental Research*, *176*, 108518. https://doi.org/10.1016/j.envres.2019.05.049

UNDRR. (2015, March). *Sendai Framework for Disaster Risk Reduction*.

Welhausen, C. (2015, November). *Visualizing a non-pandemic: Considerations for communicating public health risks in intercultural contexts*. ResearchGate. https://www.researchgate.net/publication/303731037_Visualizing_a_non-pandemic_Considerations_for_communicating_public_health_risks_in_intercultural_contexts

Witte, K. (1996). Predicting Risk Behaviors: Development and Validation of a Diagnostic Scale. *Journal of Health Communication*, *1*(4), 317–342. https://doi.org/10.1080/108107396127988

TOWARDS AN AI-ENHANCED KANSEI ENGINEERING FRAMEWORK

Emil Andersson[1], Noah Löfgren[1] Nuno Valverde[2] Lluis Marco Almagro3, , Simon Schütte[1]

[1]*Linköping University, Linköping, Sweden, simon.schutte@liu.se*

[2]*Logitech Europe SA, Switzerland*

[3]*Universitat Politècnica de Catalunya,, Spain*

ABSTRACT

Kansei Engineering (KE) is a methodology for translating subjective user emotions into tangible product design properties; however, traditional KE relies heavily on manual, labor-intensive, and subjective processes. This work explores the integration of Artificial Intelligence (AI) into the KE workflow to automate and reinforce specific KE-tasks with specialized AI-Tools. To achieve this, the research proposes a novel Multi-Agent System (MAS) framework comprising 21 specialized AI agents designed to operate across the semantic space, product property space, and synthesis phases of KE. Key components, including a Kansei Word Extractor, a Product Property Extractor, and a Synthesis Agent, were prototyped, creating a structured pipeline including language models alongside vision-language models. The framework's performance was evaluated against three historical manual KE case-studies (men's everyday shoes, battery drills, and kettles) and practically applied in a case study on Logitech gaming headphones. Findings indicate that the AI-integrated framework enables rapid ideation, successfully executing a full KE pipeline analysis in under an hour. However, empirical cross-study comparison revealed low cell-by-cell agreement with the manual case-studies. While the AI demonstrated high proficiency in bottom-up, surface-level visual extraction (e.g., shape, color, material), it fundamentally struggled to capture the holistic, non-visual, and ergonomic nuances inherent to true Kansei. The study concludes that while structured pipelines, running smaller AI models, present profound opportunities for scaling data aggregation and mitigating the "black box" phenomenon, they should accompany rather than replace human empathetic judgment.

Keywords: Kansei Engineering, Artificial Inteligence, Large Language Models, Multi-agent AI, Headphones

1 INTRODUCTION

1.1 Background and motivation

Kansei Engineering (KE) is a well-established methodology for translating human emotions and subjective perceptions into tangible product design parameters. Despite its strengths, the traditional KE process is often time-consuming, manually intensive, and reliant on subjective

interpretation. Recent advances in Artificial Intelligence (AI), particularly in large language models (LLMs) and generative systems, offer new opportunities to support and automate key steps in this process.

At the same time, the integration of AI into KE, while emerging in recent years (Liu, 2025), remains fragmented. Individual studies have started to explore AI-supported methods, there is a lack of structured understanding of how these tools map onto the established KE process.

This work is based on a master thesis work carried out at Linköping university by Andersson and Lofgren (2026).

1.2 Artificial intelligence in Product design

Artificial intelligence has become increasingly relevant in product design because it can quickly support the analysis, interpretation, and generation of design-related information. In contrast to traditional design support tools, which often rely on predefined rules or manual interpretation, AI-based methods are able to process large amounts of textual, visual, and numerical data and identify patterns that may be difficult and are time consuming to capture through conventional methods.

Several AI technologies are particularly relevant for product design. Natural Language Processing can be used to analyze textual data, such as user reviews, interviews, product descriptions, and market reports. This makes it possible to extract user needs, emotional expressions, and product-related perceptions from large text sources. Computer vision can support the analysis of visual product characteristics, such as shape, color, proportion, texture, and layout. This enables product properties to be identified and categorized more systematically from image-based material.

Generative AI further expands the role of AI in design by supporting the creation of new product concepts, visual alternatives, and design variations based on given requirements or target impressions. Large Language Models can also assist in organizing design knowledge, generating semantic descriptors, comparing alternatives, and supporting early-stage design reasoning. Together, these technologies make it possible to use AI not only as an analytical tool, but also as a generative and interpretive support throughout the product design process.

For Kansei Engineering, these developments are particularly relevant because the method depends on connecting subjective user impressions with concrete product properties. AI offers new possibilities to support this connection by extracting emotional descriptors from text, identifying visual design attributes from images, and modelling relationships between perception and design features. However, the use of AI in this context needs to be structured carefully so that individual tools are connected to the specific stages and logic of the Kansei Engineering process.

1.3 Research gap and problem formulation

Existing work demonstrates that AI can support specific KE tasks, such as Kansei word extraction or emotional prediction. However, there is limited research that systematically maps

these AI capabilities to the full KE workflow. This raises the following problem: How can AI-supported methods be systematically integrated across the Kansei Engineering process?

The aim of this paper is to analyse and structure the integration of AI into the Kansei Engineering process.

2 KANSEI ENGINEERING THE GENERAL MODEL

The Kansei Engineering process, illustrated in Figure 1, is based on the idea that a product domain can be described from two complementary perspectives: The semantic space, capturing users' emotions and perceptions, and the space of product properties, describing the physical and design attributes of the product Each of these perspectives is represented as a structured "space." In the synthesis phase, the two spaces are linked in order to identify which product properties evoke specific emotional responses. This relationship is then validated and refined through an iterative process, leading to the development of a predictive model connecting product attributes and user perception (S. Schütte et al., 2023).

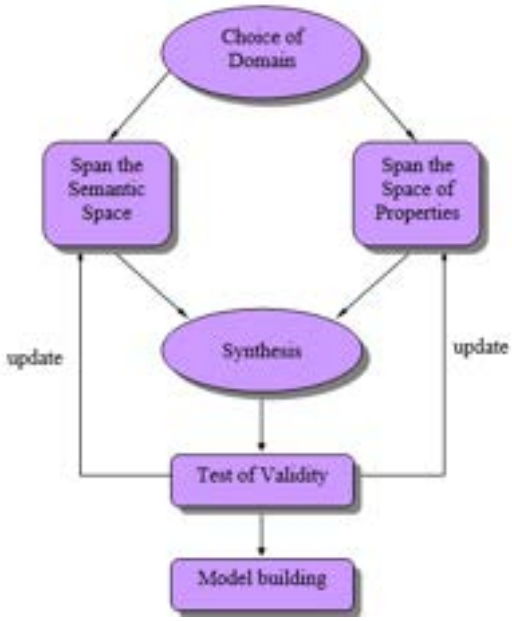


Figure 1 - The general Kansei Engineering model (S. Schütte et al., 2023; Simon Schütte et al., 2004).

2.1 Key components of the Kansei Engineering model and AI-supported extensions

The Kansei Engineering process can be understood as a sequence of structured steps that gradually transform subjective user impressions into concrete product design parameters. In the conventional process, these steps are mainly performed through manual data collection, expert interpretation, user studies, and statistical analysis. In an AI-enhanced process, the same structure can be retained, but several activities can be supported, accelerated, or partly automated through AI-based tools. This makes it possible to preserve the methodological logic of Kansei Engineering while increasing scalability, consistency, and analytical depth (see Table 1).

Table 1 - Kansei Engineering process and AI replacement models (Andersson & Löfgren, 2026).

KE Stage	Analytical Task	AI Toolbox	AI Tool developed
Semantic space	Extract and structure Kansei descriptors	NLP, LLMs	Kansei Word Extractor
Product properties	Identify and categorize design features	Computer vision	Product Property Extractor
Synthesis	Map emotions to design attributes	Machine learning, generative AI	Synthesis Agen

2.2 Choice of Domain

The process starts with the choice of domain. In the conventional Kansei Engineering process, this step involves defining the product domain, target users, market segment, and product scope. Representative product samples are then collected to cover the domain as broadly as possible. This is typically done through market analysis, expert input, competitor analysis, and the manual selection of existing products or concept examples.

AI can support this step by assisting in the identification and structuring of relevant product domains. Large Language Models and web-based AI tools can be used to summarize market trends, identify competing products, and extract relevant product categories from textual sources such as product descriptions, reviews, reports, and online catalogues. Clustering methods can also be used to group products based on similarities in descriptions, specifications, or visual appearance. In this way, AI can help create a broader and more systematic overview of the domain before the detailed Kansei analysis begins.

2.3 Spanning the Semantic Space

The semantic space represents the emotional and perceptual dimension of the product domain. In the conventional process, this step focuses on collecting Kansei words that describe how users perceive and emotionally respond to products. These words are usually gathered from literature, interviews, questionnaires, expert workshops, product reviews, magazines, advertisements, and design references. The collected words are then reduced and structured, often through the Semantic Differential Method, where users evaluate products on bipolar adjective scales such as simple–complex or elegant–rough. Statistical methods such as factor analysis, principal component analysis, or clustering are then used to identify underlying semantic dimensions.

AI can support the construction of the semantic space through Natural Language Processing and Large Language Models. NLP methods can be used to extract emotionally relevant words and expressions from large textual datasets, such as online reviews, user comments, interviews, forums, and product descriptions. Large Language Models can further support this process by generating candidate Kansei words, refining word lists, removing duplicates, grouping similar terms, and suggesting bipolar adjective pairs. Embedding-based methods can represent words as vectors, making it possible to calculate semantic similarity and cluster descriptors automatically. This allows the semantic space to be constructed from larger and more diverse data sources than would normally be feasible in a manual process.

2.4 Spanning the Space of Product Properties

In parallel with the semantic space, Kansei Engineering also requires the construction of the space of product properties. In the conventional process, this step involves identifying the physical and design-related attributes that characterize the selected product domain. These attributes may include shape, color, material, texture, size, proportions, interface elements, component layout, and other observable product characteristics. Manual methods such as expert analysis, brainstorming, affinity diagrams, Pareto analysis, and focus group evaluations are often used to collect, organize, and prioritize relevant product attributes.

AI can support this step by using computer vision and machine learning methods to identify and classify product properties from visual material. Image recognition models can analyze product images and detect relevant design features such as shape, color, texture, proportions, and component configuration. Pre-trained vision models and convolutional neural networks can be used to extract visual features, while clustering algorithms can group products or components according to visual similarity. Classification methods can also help categorize design attributes and identify dominant product characteristics across a dataset. This enables a more scalable and consistent construction of the product property space, especially when many product examples are available.

2.5 Synthesis

The synthesis step connects the semantic space with the space of product properties. In the conventional process, this means identifying relationships between Kansei words and product attributes. The goal is to understand which design features contribute to specific emotional or perceptual responses. This can be done qualitatively through expert judgement, design interpretation, and QFD-like reasoning, or quantitatively through regression analysis, conjoint analysis, neural networks, and other statistical modelling techniques.

AI can strengthen the synthesis step by modelling complex relationships between emotional descriptors and product attributes. Machine learning models can be trained to predict user perceptions based on product features or to identify which features are most strongly associated with specific Kansei words. Regression models, decision trees, random forests, support vector machines, and neural networks can be used depending on the type and amount of data available. In addition, generative AI models can be used to translate semantic targets into new design concepts. For example, image generation or diffusion models can generate product concepts intended to express Kansei words such as elegant, robust, playful, or comfortable. In this way, the synthesis step can move beyond analysis and become a generative design activity.

2.6 Test of Validity

The test of validity examines whether the Kansei model accurately represents the relationship between user perception and product properties. Conventionally, this may involve comparing predicted and observed user responses, conducting additional user evaluations, and revising the semantic space, product property space, or synthesis model.

AI can support this step by testing model performance, identifying inconsistencies, and comparing new user feedback with predicted Kansei responses. However, because emotional perception is context-dependent, human evaluation remains necessary. AI should therefore support, rather than replace, human judgement in validity testing.

2.6.1 Model Building

In the final step, the results of the Kansei Engineering process are combined into a model that describes how product properties influence user perception. Conventionally, this model may take the form of design guidelines or a predictive model linking product attributes to Kansei responses.

In an AI-enhanced process, model building can instead become an integrated design-support workflow that combines language models, computer vision, machine learning, and generative AI. Such a system can support the extraction of Kansei descriptors, identification of product properties, modelling of emotion-design relationships, and generation of design alternatives. However, human interpretation remains essential to evaluate the relevance and meaning of the AI-generated outputs.

3 PROOF OF CONCEPT

To demonstrate the practical applicability of the proposed AI-enhanced Kansei Engineering (KE) framework, a proof-of-concept study was conducted. The study applies the AI toolbox introduced in Chapter 3 to a real product domain, allowing an evaluation of how effectively AI can support different stages of the KE process. The models were tested, trained and validated with previous published KE case studies, in the design of power drills (Grimsaeth et al., 2016), mens' shoes (Van Lottum et al., 2006) and kettles (Mamaghani & Tajoddini, 2010). Upon validating and understanding the usefulness of the developed multi-agent AI pipeline (KE-MAS), a new and more contemporary dataset emerged with gaming headphones from Logitech and its competitors.

The purpose of this proof of concept is to apply the AI-supported methods described in Chapter 3 within a complete Kansei Engineering workflow. The intention is to demonstrate the potential of AI tools to replace or augment traditional KE methods in the generation of the semantic space, the definition of product properties, and the synthesis between these domains. Rather than aiming for full validation of performance, the study focuses on demonstrating feasibility and exploring how AI can support the execution of a full KE pipeline.

3.1 Materials

The image material consisted of 15 commercially available gaming headphones, samples selected to represent visual variation within the product domain. These are shown in Figure 2. The samples were retrieved from online retailers and included headphones displaying different designs and brands. This allowed the Product Property Extractor to identify recurring components and design variants across the domain, such as the headband, ear cushions, ear cups, and microphone boom. The samples were therefore not selected to represent a statistically complete market overview, but to offer a testbed for the AI-supported Kansei Engineering pipeline.



Figure 2 - A dataset of 15 gaming headphones were used for this study (Andersson & Löfgren, 2026).

3.2 Method

The study applied the developed multi-agent AI pipeline (KE-MAS) across the Kansei process (Andersson & Löfgren, 2026). This allowed a full Kansei pipeline execution without direct involvement of human participants.

3.2.1 Step 1: Kansei Word Extraction

In the first step, textual user data were analyzed to identify Kansei words describing users' emotional and perceptual responses to gaming headphones. Different text sources were considered, including product-specific reviews, general headphone reviews, and UX data from Logitech. The purpose was to generate a semantic space that reflected how users describe and experience the product domain.

3.2.2 Step 2: Product Property Extraction

In the second step, product images were analyzed to identify visible design components and product properties. The AI system decomposed the headphones into components such as headband, ear cushions, ear cups, and microphone boom. For each component, design variants related to shape, material, color, articulation, and ornamentation were identified.

3.2.3 Step 3: Application of the Synthesis Agent

In the final step, the extracted Kansei words were connected to the identified product properties. The Synthesis Agent evaluated how different design variants were associated with selected Kansei words and generated design configurations for target impressions such as appealing, elegant, and comfortable. These configurations were then used to create visual product concepts, demonstrating how AI can support the full Kansei workflow from user perception to design output.

4 RESULTS

The results are presented according to the three main steps of the AI-supported Kansei Engineering pipeline: Kansei Word Extraction, Product Property Extraction, and Synthesis.

4.1 Step 1: Results from Kansei Word Extraction

In the first step, three types of textual data were compared: product-specific reviews, general headphone reviews, and UX data from Logitech. The UX dataset produced the most relevant

semantic space for the proof of concept, since it contained more product-specific and design-relevant descriptions of user perception. In contrast, the general review data included a larger amount of functional and post-use evaluations, such as comments related to sound quality, technical performance, or long-term use.

The extracted Kansei words from the UX data included descriptors such as comfortable, appealing, lightweight, elegant, soft, stiff, intuitive, ergonomic, stable, and solid. These words describe both visual and tactile impressions of the headphones and were therefore suitable for the following synthesis step. The result indicates that the quality and focus of the input data strongly influence the semantic space generated by the AI pipeline.

4.2 Step 2: Results from Product Property Extraction

In the second step, the AI system analyzed the headphone image dataset and decomposed the products into four main components: headband, ear cushion, ear cup, and microphone boom. For each component, the system identified visible design properties such as shape, material, color, articulation, and ornamentation.

Across these components, the system identified 22 design variants. Examples include curved or angular shapes, soft or rigid materials, black or vibrant colors, flexible or fixed articulation, and simple or colorful ornamentation. The result shows that the AI system was able to generate a structured product property space based on visual information from the image dataset. This demonstrates how computer vision and multimodal AI can support a task that is traditionally performed manually through expert analysis.

4.3 Step 3: Results from the Synthesis Agent

In the final step, the extracted Kansei words were connected to the identified product properties. The Synthesis Agent evaluated how different design variants contributed to selected Kansei impressions and generated design configurations for target words such as appealing, elegant, and comfortable.

The configuration for appealing emphasized expressive and visually dynamic features, such as vibrant colors, colorful ornamentation, and a more gaming-oriented visual appearance. The configuration for elegant was characterized by restrained design choices, including black colors, simple ornamentation, curved shapes, and a more minimal visual expression. The configuration for comfortable emphasized soft materials, rounded forms, flexible structures, and deep ear cushions.

These results were translated into visual headphone concepts Figure 3. The generated concepts were distinct and coherent in relation to the target Kansei words, suggesting that the AI pipeline can support the movement from user perception to product property analysis and further to early design concept generation. At the same time, the synthesis was strongest for visually inferable impressions, while qualities that depend on long-term use or technical performance remain more difficult to assess through image-based AI analysis alone.



Figure 3 - Rendering of headphones corresponding to certain feelings and emotional descriptors: left: “appealing headphones”, center: “elegant” headphones, and right: “comfortable” headphones (Andersson & Löfgren, 2026).

5 DISCUSSION

The proof-of-concept study shows that AI can support the main stages of the Kansei Engineering process, from extracting Kansei words to identifying product properties and generating design concepts. The results suggest that AI can make the process more efficient and systematic, especially in the early phases of analysis and ideation.

However, the study also shows that the quality of the results depends strongly on the input data. The UX dataset produced more relevant Kansei words than general online reviews, which indicates that AI-supported KE requires carefully selected and context-specific data.

The pipeline performed best for Kansei impressions that can be visually inferred, such as appealing, elegant, and comfortable. Qualities related to long-term use, sound performance, technical reliability, or deeper ergonomic experience were more difficult to assess through image-based AI analysis alone.

Overall, the results indicate that AI should be seen as a support tool rather than a replacement for human expertise. The generated concepts can serve as useful starting points for designers, but they still require interpretation, refinement, and validation by humans.

6 CONCLUSIONS

The main conclusion from this work is that AI can be integrated into the established Kansei Engineering structure without replacing the overall model as Figure 4 shows. The contribution is therefore not a new Kansei Engineering process, but an AI-supported extension of the existing one. Figure 4 illustrates how AI tools can be positioned within the traditional workflow to support selected stages while preserving the basic logic of the method.

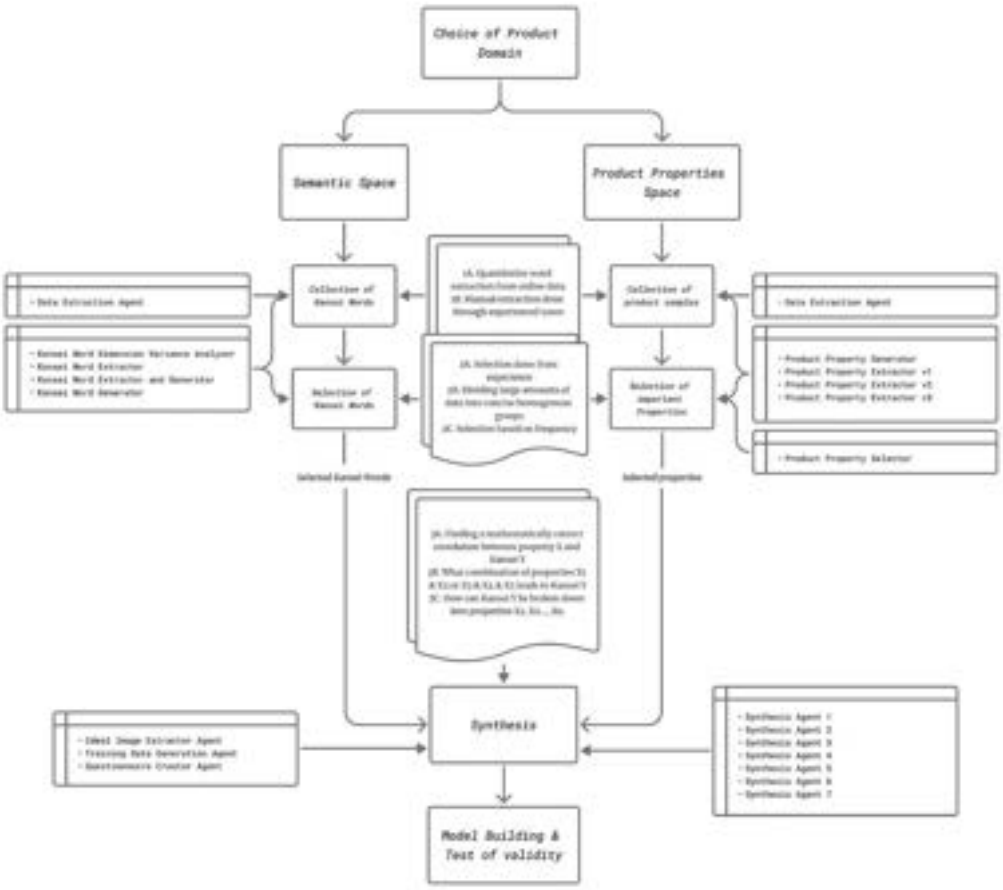


Figure 4 - The general Kansei Engineering model enhanced with AI tools.

AI should however not be treated as one general-purpose tool in Kansei Engineering. Different AI technologies support different tasks. NLP and large language models are suitable for extracting and structuring Kansei words, computer vision can support the identification of product properties, and generative AI can be used to support synthesis and concept visualization. This means that AI integration in KE should be understood as a toolbox, where different tools are matched to specific methodological tasks.

The framework also shows that AI is especially valuable in the early and resource-intensive stages of Kansei Engineering. Collecting user expressions, analyzing large amounts of text, identifying product properties from images, and generating early design directions are all tasks where AI can reduce manual effort and increase speed. The proof of concept with gaming headphones confirms that the proposed structure can be applied in practice, moving from user data to product properties and finally to visual concept generation.

At the same time, the integration of AI introduces important risks. Since AI can automate or simulate many steps of the process, including user-like ratings, there is a risk that human interaction is removed too far from a method that is fundamentally intended to support customer-focused product development. Just because AI can replace a certain activity does not

mean that it should. In Kansei Engineering, human judgement and real user input remain essential for ensuring that the results reflect meaningful emotional responses rather than only AI-generated approximations.

A final conclusion is that AI-generated results can be difficult to question because they often appear coherent and convincing. This creates a risk that incorrect or weakly grounded outputs are mistaken for valid results. Therefore, AI-supported Kansei Engineering should include critical human review, transparency in how results are generated, and validation with users or experts. AI should be used as a structured support system, not as an autonomous replacement for the human-centered foundation of Kansei Engineering.

7 ACKNOWLEDGMENTS

The authors would like to express their appreciation to Logitech Europe SA., for the continuous and extensive support to this research.

8 REFERENCES

Andersson, E., & Löfgren, N. (2026). *Evaluation and Integration of AI in the Kansei Engineering Process*. Linköping University.

Grimsaeth, K., Gangwar, M., Baggerud, B., & Narain, R. (2016). Kansei Engineering Based Emotional Product Design: An Illustration by a Battery Drill Case-Study. *Kansei Engineering International Journal*, *9*(2), 67–79. https://doi.org/10.5057/kei.9.67

Liu, J. (2025). AI in Automated and Remote UX Evaluation: A Systematic Review (2014–2024). *Advances in Human-Computer Interaction*, *2025*(1). https://doi.org/10.1155/ahci/7442179

Mamaghani, N. K., & Tajoddini, M. (2010). Design with emotional approach by implementing kansei engineering-case study: design of kettle. *International Conference on Kansei Engineering and Emotion Research*, 8.

Schütte, S., Mohd-Lokman, A., & Marco-Almagro, L. (2023). *The general Kansei Engineering model*.

Schütte, Simon, Eklund, J., Axelsson, J. R. C., & Nagamachi, M. (2004). Concepts, methods and tools in Kansei engineering. *Theoretical Issues in Ergonomics Science*, *5*(3), 214–231. https://doi.org/10.1080/1463922021000049980

Van Lottum, C., Pearc, K., & Coleman, S. (2006). Features of Kansei engineering characterizing its use in two studies: Men's everyday footwear and historic footwear. *Quality and Reliability Engineering International*, *22*(6), 629–650. https://doi.org/10.1002/qre.803

UNIVERSAL PERCEPTION, CULTURAL INTERPRETATION
A CROSS-CULTURAL COMPARATIVE UX STUDY OF COLOUR PERCEPTION IN AUSTRALIA AND INDIA

Armaan Kumar

Delhi Technological University, India, armaankumar6128@gmail.com

ABSTRACT

This study investigates the strength of correlation between colour categories and emotional or behavioural responses across cultural contexts, focusing on primary (red, blue, yellow) versus secondary colours (green, orange, purple) in populations from India and Australia. Using parallel online surveys, participants from both countries reported their emotional associations and cultural symbolism for each colour. Analysis demonstrates that primary colours have a higher degree of cross-cultural alignment than secondary colours, indicating stronger correlational stability. In contrast, secondary colours exhibit greater variation in meaning, shaped by culturally specific interpretations. These results support the hypothesis that while both colour categories evoke shared emotional responses, the correlation between primary colours and their associated emotions is more robust across national contexts. The findings offer practical guidance for User Experience (UX) designers in global and culturally sensitive colour system design.

Keywords: Primary colours, Secondary colours, Correlational Research, Cultural Relativity, Proportional Divergence

1 INTRODUCTION

This research paper essentially tries to understand the correlational value between primary colours and the universality of the emotional and/or behavioural responses they elicit from people across cultures and similarly for the secondary colours. In today's times when globalisation is leading to universality in almost every sector and industry, finding cultural commonality rooted in biology becomes very important for User Experience (UX) Research.

Colours open an accessible gateway to universal design that can transcend cultural barriers especially in the context of UX. This paper attempts to assess whether primary colours offer a higher consistency in their associated meanings or connotations across cultures as compared to secondary colours making our CORE RESEARCH QUESTION the following: -

"Is there a higher correlational value between primary colours and the emotional or behavioral responses they elicit as compared to secondary colours?"

1.1 Objectives

- Firstly, to analyse whether primary colours and their associated meanings can help us shape the future of Universal Human Design.
- Secondly, to understand how cultural variations in psychological appraisals by individuals affect the way different colours represent different emotions.
- Lastly, to provide UX design reccommendations on when to use colours with universal connotations and when to localize colour choices for culturally diverse users.

1.2 Scope of Research

The main focus is to understand how culture and visual context affect perception of primary and secondary colours and in doing so, the scope spans over the following fields: -

1. Psychology

The research paper heavily incorporates a study of behavioural science and emotional responses to visuals of a colour. It tries to understand how common are the meanings associated with certain colours in different cultural and geographical contexts. The study has also been conducted keeping in mind the biological and societal influences of colour perception on the human mind. However, the paper does not cover any clinical psychological implications like colour therapy, etc.

2. Cognitive Science and Biology

As stated earlier, biology does play an important role when it comes to colour perception. For example, red has always been known to have a warning effect on the human mind due to multifarious reasons like its natural association with fire, blood or injury (Science Focus, n.d.) (Mather, 2020). Moreover, red has the highest wavelength of all the colours meaning it can be seen over greater distances as well. Secondly, the L-cones or the red receptors in our retina are located more centrally in the fovea of the eye, making red a more visually striking colour (Bosten & Boehm, 2015).

Furthermore, evolutionarily, bright colours are associated with danger like in the coloration of animals such as certain poisonous dart frogs, the blue ringed octopus, etc which use aposematism as a predator-warning system of their bad taste and toxicity (Bosten & Boehm, 2015). However, the study is limited to general cognitive-emotional reactions, not neurological imaging or medical-level biological analysis.

3. Cultural Studies

The paper also examines how cultural context influences colour meaning, symbolism, and emotional associations. Societal elements like history, economic status and access due to power have historically affected the way colour is perceived. For instance, the general association of purple to luxury or royalty comes from the fact that in ancient times, even before the renaissance, the colour purple was available in its natural form only as the Tyrian purple extracted from the glands of the Murex snails. The process of

extraction was so difficult that only the rich could afford to use the colour and thus with time it was solely associated with them (Refashioning Renaissance, 2023).

Thus, the paper tries to analyse how cultural context and use of a colour in a certain way can affect the way it is perceived because it is generally agreed upon that while biology sets certain fixed constraints, the environment of an individual determines several of their characteristics (including their perception) within those constraints. However, the paper focuses on comparative patterns across geographical regions rather than deep ethnographic studies of individual cultures.

4. Design Studies

The findings of this paper attempt to stretch across the scope of several design domains including visual design, user experience and interaction design. Essentially trying to understand how colour universality or cultural specificity should inform interface design, branding, and user engagement strategies. Furthermore, it aims to identify when designers should rely on colours with universal emotional connotations versus when local cultural adaptations are required. The scope in this field is though limited to interface design and branding, not material colour science in industrial design.

2 LITERATURE REVIEW

Studies in psychology, anthropology, and design show that colours influence emotion and behaviour, but the consistency of these reactions varies (Ou, Luo, Woodcock, & Wright, 2004). Primary and secondary colours both show stable interpretations across cultures because they correspond to distinct wavelengths that the human eye processes with greater clarity, while also carrying more culturally shaped meanings since they are also learned through social and historical associations. This provides the foundation for examining the degrees to which primary colours and secondary colours elicit universal responses each and then finding the difference between them to prepare a comparative study, particularly for global UX design.

2.1 Existing Theories on Colour Perceptions

1. Basic Colour Term Theory (Berlin & Kay, 1969)

This theory states that certain colours, especially primary ones, appear consistently across languages. The authors describe these colours as having strong focal clarity, supporting the idea of universal emotional responses. In essence, first there is a separation of light and dark, then red, then yellow or green and finally blue. This order is consistent due to the fact that biological human visual systems (cones in the retina, etc) are similar across cultures. This indicates that linguistically at least, primary colours seem to have a greater consistency across cultures as compared to other colours.

2. Opponent Process Theory (Hurvich & Jameson, 1957)

This biological model explains that the visual system interprets colour in opposing pairs such as red versus green and blue versus yellow. Since these processes are physiological, interpretations of basic colours show some universal tendencies.

3. Ecological Valence Theory (Palmer & Schloss, 2010)

This theory argues that people prefer colours based on the objects they associate them with. Since objects differ across cultures, secondary and complex colours vary more in meaning.

4. Universalists Vs. Relativists Theories (Adams & Osgood, 1973)

Universalists argue that colour meanings arise from biological perception, while relativists argue that meaning is culturally constructed. This distinction is central to comparing primary and secondary colour interpretations.

2.2 Gaps in Current Research and Theories

There are several gaps in the current theories and research that this paper attempts to bridge. Firstly, there are no studies that deal with primary and secondary colours separately and attempt to draw a comparative analysis between the two. This research paper, on the other hand, is clearly defining the two as separate variables to draw on their predictability for design.

Secondly, theories like the EVT, OPT and BCTT all explore HOW colour is perceived across cultures but relational studies regarding colour perceptions still lack depth and have not been touched upon. This paper however, clearly tries to quantitatively analyse the relativity or universality of primary and secondary colours separately.

Thirdly, previous theories are based highly on outdated contexts and are West focussed. This paper however, takes India and Australia as the two main focal points, trying to study a greater cultural diversity and taking away the focus from the already deeply studied west.

3 PROPOSED HYPOTHESES

Keeping the literature review in mind, this paper attempts to prove the following hypotheses: -

3.1 Hypothesis 1

The correlation between primary colours and the emotional or behavioural responses they elicit will be stronger across cultures than the correlation observed for secondary colours.

3.2 Hypothesis 2

Both primary and secondary colours will show some degree of shared emotional associations across cultures, but primary colours will demonstrate a higher level of consistency in these shared responses.

3.3 Hypothesis 3

Participants from different cultural backgrounds will exhibit more similar emotional interpretations for colours regardless of whether they are primary or secondary, indicating a stronger cross-cultural alignment for hues.

3.4 Hypothesis 4

Variations in responses to secondary colours will be greater than variations in responses to primary colours, resulting in a lower correlational strength for secondary colours across cultural groups.

4 METHODOLOGY

This paper uses a correlational research design to examine the strength of the relationship between colour categories and the emotional or behavioural responses they elicit across two cultural contexts. Correlational research is appropriate because the variables are not manipulated.

Instead, naturally occurring responses from participants are measured to determine whether the relationships between colours and emotional interpretations are stronger for primary colours than for secondary colours. The aim is to calculate the degree of association between the two variables in each colour category and compare the correlational strength across cultures. For a hypothetical example,

Blue shows the highest association with “X” emotion.
What percentage of Indians agree with the statement?
What percentage of Brits/ Australians agree with the statement?
What is the proportional divergence/similarity?

Orange shows the highest association with “X” emotion.
What percentage of Indians agree with the statement?
What percentage of Brits/ Australians or Brits agree with the statement?
What is the proportional divergence/similarity?

Then, we compare the two results and try to understand whether the primary colour affords a higher correlation between the two elements or a secondary colour.

4.1 Sample Size

Two independent samples were collected, one from India and one from Australia. The sample size for each one was 20 participants making the total sample size 40. Each sample completed an identical Google Form that measured emotional and cultural responses to both primary and secondary colours. The dual-form approach ensured cultural specificity while maintaining question consistency for accurate comparison. Participants represented a general population sample and were recruited organically through personal networks and online sharing.

4.2 Tools Used

Both primary and secondary colours will show some degree of shared emotional associations across cultures, but primary colours will demonstrate a higher level of consistency in these shared responses. Two Google Forms were designed with identical structures and questions. Each form included four sections: -

1. Emotional responses to primary colours,
2. Emotional responses to secondary colours,
3. Use of colour in UX design,
4. Cultural and symbolic associations with colours.

All questions were closed-ended (MCQ or MSQ), in line with correlational methodology and quantitative analysis requirements. Some questions included colour-based image prompts to standardise visual perception across respondents.

4.3 Variables

There are two variables in the study, the dependent and the independent variable: -

1. Independent Vatiable: Colour Category (primary or secondary colours)
2. Dependent Variable: Emotional responses (e.g., calm, happiness, confusion, alertness, sadness, energy). Behavioural interpretations (e.g., danger, celebration, caution). Cultural and symbolic associations (e.g., mourning, auspiciousness, positivity).

The purpose is to determine the correlational strength between each colour category and the consistency of responses across cultures.

4.4 Data Analysis

Correlational analysis forms the core method of interpretation. The study examines: -

- The degree of similarity between Indian and Australian responses for each primary colour and secondary colour.
- The consistency of emotional associations within each culture for both colour categories.
- The comparative correlational strength between primary colour responses and secondary colour responses.

Although statistical correlation coefficients (r values) have not be generated numerically due to categorical variables and responses rather than numerical ones, conceptual correlation has been derived by examining response frequencies, proportional similarity, and cross-cultural alignment in selected options. Patterns that repeat strongly across both cultural samples indicate higher correlational strength.

5 RESULTS AND ANALYSIS

This paper now presents the comprehensive findings from the cross-cultural survey administered to participants in India and Australia, followed by a quantitative analysis of the data to test the four proposed hypotheses regarding the universality and cultural relativity of colour perception.

5.1 Sample Demographics

The study collected a total of N=40 complete responses, comprising two independent samples of N=20 participants, each from India and Australia. This balanced sample size allows for a direct proportional comparison between the two distinct cultural groups.

5.2 Cross-Cultural Consistency in Primary Colour Perception (Hypotheses 1.2 and 3)

Analysis of responses related to primary colors (Red, Blue, Yellow) showed a significantly high degree of proportional similarity in both emotional and functional associations, supporting the biological and universal roots of primary color meaning.

Table 1. Analysis of Primary Colours

Primary Colour	Most Associated Emotion	India (N=20)	Australia (N=20)	Proportional Similarity
Red	Anger	11 (55.0%)	13 (65.0%)	High
Blue	Sadness	9 (45.0%)	8 (40%)	High
Yellow	Happiness	9 (45.0%)	9 (45.0%)	Perfect

Yellow: The association of Yellow with Happiness was statistically identical in both samples (45.0% each).

Red & Blue: The leading emotional responses were consistent, with minimal proportional difference (less than 10 percentage points).

Functional Alignment (Positive and Trust)

Consistency extended into functional domains related to User Experience (UX) and branding, further reinforcing the universal interpretation of primary colors (supporting Hypothesis 2).

Table 2. Analysis of Functional Alignment

Functional Question	Top Response	India (N=20)	Australia (N=20)	Proportional Alignment
Most universally positive	Yellow	19 (95.0%)	14 (70.0%)	Strong Alignment
Trust Most in a Brand logo	Blue	12 (60.0%)	11 (55.0%)	Strong Alignment

Universal Positivity: Yellow was overwhelmingly selected by both groups as the most universally "positive" colour, with a remarkable 95.0% consensus in India.

Trustworthiness: Blue was the top choice for the colour to trust MOST in a brand logo for both groups (60.0% India vs. 55.0% Australia), indicating a globally shared connotation of reliability and security for this hue in UX design, thus supporting the higher consistency for primary colours (H1).

5.3 Cross-Cultural Variability in Secondary Colour Perception (Hypothesis 4)

Responses for secondary colours (Orange, Green, Purple) showed significantly greater proportional divergence when analyzing general and culturally specific associations, indicating that secondary hues are more heavily influenced by contextual and historical factors (supporting H4).

General and Cultural Association Divergence (Green, Purple, Orange)

The consistency for secondary colours was markedly lower than for primary colours across all metrics. The general association for Green showed the largest proportional divergence.

Table 3. Analysis of Secondary Colours

Secondary Colour	Top Response	India (N=20)	Australia (N=20)	Proportional Divergence
Green	Nature	11 (55.0%)	5 (25.0%)	High (30.0% difference)
Purple	Royalty/ Luxury	15 (75.0%)	10 (50.0%)	High (25.0% difference)
Orange	Religious/ Sacredness	6 (30.0%)	4 (20.0%)	Low (10% difference)

Green: The association of Nature with green is highest in both the countries but there is a high difference (30%) in the percentage of people agreeing with it with 25% of Australians and 55% Indians supporting Hypothesis 1 and 2.

Purple: Similarly, the association of Royalty/ Luxury with Purple is highest in both the countries but there is a high difference (25%) in the percentage of people agreeing with it with 50% of Australians and 75% Indians supporting Hypothesis 1 and 2.

Orange: However, in the case of Orange, while the divergence is low (10%) and the highest voted emotion is religious/sacredness across cultures, the overall response for the orange is low and below 50% on both surveys (30% and 20%) in general. This indicates lower agreement on the meaning of orange in both cultures.

5.4 Reversing Emotion-Colour pair for UX Analysis

In this section, rather than giving colours and asking for the emotions associated with it, the survey presents emotions in the context of UX choices that better help in understanding colours associated with certain emotions.

Table 1. UX Analysis

UX Metric	Top Response	India (N=20)	Australia (N=20)
Banking App Interface	Blue	14 (70.0%)	5 (25.0%)
Health and Wellness	Green	13 (65.0%)	9 (45.0%)
Attention Grab	Red	14 (70.0%)	6 (30.0%)
Luxury or Sophistication	Golden (India) Black (Australia)	Golden (9, 45.0%)	Black (7, 35%)

Agreement: Blue, Green and Red all seem to be the top choice for Trust (in banking interfaces), Health and Wellness and Attention-Grabbing (in website banners), respectively. This shows a level of agreement cross culturally regarding the colours associated with certain emotions supporting regardless of them being secondary or primary, limitingly supporting Hypothesis 2.

Divergence: Luxury and Sophistication however, are completely different in the two cultures being represented by golden in India and Black in Australia, demonstrating cultural influence on colour perception supporting Hypothesis 3.

5.5 Cultural and Symbolic Colour Analysis

When it comes to cultural and symbolic questions, they demonstrated the greatest cross-cultural divergence, particularly for ritualistic or emotional concepts like mourning and auspiciousness.

Table 1. Cultural Symbolism Analysis

Cultural Metric	India's Top Response	Australia's Top Response	Proportional Divergence
Auspicious or Lucky	Yellow (6, 30.0%)	Red (5, 25.0%)	Complete Divergence
Mourning or Loss	White (8, 40.0%)	Black (11, 55.0%)	Complete Divergence
Calmness or Peace	White (9, 45.0%)	White (11, 55.0%)	Low (10.0% difference)

Auspiciousness: Different colours are highly associated with auspiciousness in the two countries, Yellow in India and Red in Australia. It can be noted that both the colours are primary.

Mourning: There is diametric divergence when it comes to mourning or loss where in India, White is associated with the concept the most and Australia it is Black.

Peace: However, in the case of peace, there is a higher symbolic similarity and the divergence is low. Both cultures associate white with it the most, suggesting that the achromatic colour's purity and simplicity holds a universally positive association with tranquility.

6 DISCUSSION

This cross-cultural study comparing color perceptions in India and Australia supports a hierarchical model of color meaning: universal cognitive mechanisms dominate primary hues, while cultural relativity shapes secondary, symbolic, and application-specific meanings. These findings refine debates in color psychology and provide practical guidance for global design.

6.1 Contributions to Academic Colour Theory

This research provides quantitative evidence that helps reconcile the universalist–relativist debate in color psychology by identifying where universality holds and where cultural influence dominates.

- Refining Universalism (Primary Colours): Very low divergence (0–10%) in primary color responses supports Basic Colour Term Theory (BCTT) (Berlin & Kay, 1969), confirming that cognitively fundamental hues like red, yellow, and blue show strong cross-cultural stability.

- Relativity Begins with Secondary Hues: Divergence increases significantly for secondary colors (e.g., Green/Nature at 30%), indicating that cultural influence emerges beyond the primary set, where meanings are more open to historical and environmental shaping.

- Cultural Inversion in High-Context Domains: Symbolic and application-specific colors (e.g., mourning, banking trust) show extreme divergence, including complete

oppositions (White vs. Black for mourning), demonstrating that cultural coding can override universal perceptual tendencies.

6.2 Contribution to Practitioner Knowledge and Global Design

The findings offer actionable insights for practitioners in branding, marketing, UX, and interface design across culturally distinct markets.

- **Strategy 1: Universal Cores, Local Shells:** Primary colours provide a stable, low-risk foundation for cross-cultural communication. High shared consensus (e.g., Yellow for positivity, Blue for trust in logos) confirms their value as global functional anchors.

- **Strategy 2: Navigating High-Risk Symbolic Divergence:** Symbolic colours cannot be used interchangeably. Complete inversions (e.g., mourning colours) and EVT-based splits (Palmer & Schloss, 2010) in luxury perception (Golden in India vs. Black in Australia) show that high-context and prestige applications require careful cultural alignment.

- **Strategy 3: The Need for Contextual Certainty:** Differences in consensus levels (e.g., stronger functional colour coding in India vs. lower standardization in Australia) indicate that some markets demand stricter adherence to established colour codes, while others allow greater flexibility.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to everyone who contributed to the completion of this research project. First and foremost, I am deeply thankful to all the participants from India and Australia who generously offered their time and perspectives. Their responses formed the core data of this study, and without their willingness to engage, this cross-cultural exploration of colour perception would not have been possible. A token of my gratitude to all my professors who assisted me with the research as well.

REFERENCES

Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology, 4*(2), 135–156. https://doi.org/10.1177/002202217300400201

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of Chicago Press.

Bosten, J. M., & Boehm, E. (2015). Neural correlates of colour processing in human visual cortex. *Frontiers in Psychology, 6*, Article 421. https://doi.org/10.3389/fpsyg.2015.00421

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64*(6), 384–404. https://doi.org/10.1037/h0041403

Mather, G. (2020, March 12). The effect of the colour red on brain waves. *Max Planck Institute for Biological Cybernetics*. https://maxplanckneuroscience.org/the-effect-of-the-color-red-on-brain-waves/

Ou, L.-C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application, 29*(3), 232–240. https://doi.org/10.1002/col.20005

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences, 107*(19), 8877–8882. https://doi.org/10.1073/pnas.0906172107

Refashioning Renaissance. (n.d.). *Shell-purple dye experiments*. https://refashioningrenaissance.eu/experiments/shell-purple-dye/

Science Focus. (n.d.). *Why are warning signs red?* https://www.sciencefocus.com/the-human-body/why-are-warning-signs-red